ABSTRACT

Title of Dissertation: ORTHOLOGOUS GENE SWAPPING AND EXPERIMENTAL EVOLUTION PROVIDE NOVEL WAY TO STUDY ESSENTIAL POXVIRUS GENES

Carey A. Stuart, Doctor of Philosophy, 2018

Dissertation directed by: Dr. Bernard Moss, Adjunct Professor, National Institutes of Health and Dr. Jeffrey DeStefano, Professor, Biological Sciences Program

The transcriptional program of poxviruses is divided into early, intermediate and late phases enabled by a multisubunit DNA-dependent RNA polymerase and stage-specific transcription factors that recognize cognate promoters. Although promoter sequences are highly conserved among the different chordopoxvirus genera, the transcription factors exhibit considerable amino acid divergence that parallels the evolutionary distance of the host species. Thus, the large/small subunits of the intermediate transcription factors (ITFs) of salmon gill poxvirus, crocodilepox, canarypox, and myxoma have 23/29, 40/31, 51/38 and 58/65 % amino acid identity, respectively, to the vaccinia virus (VACV) orthologs. The purpose of the present study was to determine the functional interchangeability of the ITF subunits and their putative interactions with other elements of the transcriptional machinery. A

quantitative readout of ITF function using firefly luciferase (Fluc) was obtained. The activity of the large subunit orthologs was greater than that of the small subunit orthologs, with both sets following the degree of sequence similarity in relation to VACV. The same pattern was obtained with both heterospecific (e.g., myxoma large and VACV small subunits) and homospecific (e.g., myxoma large and small subunits) pairings, suggesting inefficient interactions with other elements of the transcription system. When recombinant hybrid VACV expressing the Myxoma virus (MYXV) ortholog of the small subunit (A8) were blind passaged multiple times, their replicative abilities were enhanced. Complete genome sequencing of the virus populations revealed five mutations present in the two largest subunits of the viral RNA polymerase (RNAP) and two predicted expression-enhancing mutations around the translation initiation site of the MYXV A8 ortholog. Amplicon sequencing was used to quantify the frequency of each mutation in its respective population, which revealed that they increased as passaging occurred. This indicated a correlation with increased fitness, which then needed to be confirmed, so these mutations were all experimentally introduced into the original hybrid virus and demonstrated to enhance virus replication independently. These mutations were then characterized to determine their specific effects on the viral RNAP (vRNAP) and viral replication and transcription. This approach could have broader applications for studying essential genes in poxviruses and other viruses as well.

ORTHOLOGOUS GENE SWAPPING AND EXPERIMENTAL EVOLUTION PROVIDE NOVEL WAY TO STUDY ESSENTIAL POXVIRUS GENES

by

Carey A. Stuart

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Jeffrey DeStefano, Chair
Dr. Bernard Moss, Co-Chair
Assistant Professor George Belov
Professor James Culver
Dr. Alison McBride

Dedication

To my parents, Colin and Marcia Stuart, to whom I am eternally grateful for everything you have given me, and to Caspar, my cousins, and the rest of my family, for always supporting me.

Acknowledgements

I would like to thank my PI, Dr. Bernard Moss, for his seemingly limitless patience mentoring me. I would also like to thank the members of my committee for their guidance during my thesis. In addition, I would like to extend special thanks to Dr. Linda Wyatt, Catherine Cotter, Dr. Sara Reynolds, Dr. Seong-In Hyun, Dr. Tatiana Senkevich, Erik Zhivkoplias, and The Science Wives, who gave me an immeasurable amount of technical advice and moral support over the years. Finally, to all the past and current members of the Moss and Berger labs, past and current LVD Administrative Staff, and past and current people in the UMD BISI and Virology Programs, I say thank you for all your help and support over these last few years.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

aa	Amino acid
ASFV	Asfavirus
ATI	A-Type inclusion
ATP	Adenosine triphosphate
bp	Base pairs
CEV	Cell-associated enveloped virion
CFR	Case fatality rate
ChPV	Chordopoxvirus
CMLV	Camelpox virus
CMV	Cytomegalovirus
CNPV	Canarypox virus
CNV	Copy number variant or variation
CPXV	Cowpox virus
CPXV-BR	Cowpox virus Brighton Red strain
CRV	Nile Crocodilepox virus
ddPCR	Digital droplet PCR
DdRP	DNA-dependent RNA polymerase
DNA	Deoxyribonucleic acid
DNAP	DNA polymerase
dsDNA	Double-stranded deoxyribonucleic acid
DsRed	Discosoma sp. Red fluorescent protein
dsRNA	Double-stranded ribonucleic acid

ECTV	Ectromelia virus
EDTA	Ethylenediaminetetraacetic acid
EEV	Extracellular enveloped virion
EFC	Entry-fusion complex
eIF2α	Eukaryotic initiation factor 2α
EM	Electron microscopy
ER	Endoplasmic reticulum
ERGIC	Intermediate compartment between the endoplasmic reticulum and Golgi apparatus
ESM	Early stop mutation
ET	Electron tomography
EV	Enveloped virion
Fluc	Firefly luciferase
FWPV	Fowlpox virus
GFP	Green fluorescent protein
GV	Giant Virus
h	Hours
HA	Hemagglutinin
HCV	Hepatitis C virus
HGT	Horizontal gene transfer
HIV	Human immunodeficiency virus
HJ	Holliday junction
hnRNP	Heterogeneous nuclear ribonucleoprotein
hpi	Hours post-infection

HSPV	Horsepox virus
HSV	Herpes simplex virus
IAV	Influenza A virus
IBT	Isatin-β-thiosemicarbazone
IEV	Intracellular enveloped virion
IFN	Interferon
IL	Interleukin
IMV	Intracellular mature virion
Indel	Insertion/Deletion
ITF	Intermediate transcription factor
ITR	Inverted terminal repeat
IV	Immature virion
IV-L	IV-like
JRC	Jelly-roll capsid
kbp	Kilobase pairs
kDa or kD	KiloDaltons
KLC	Kinesin light chain
KSHV	Kaposi's sarcoma-associated herpesvirus
LGT	Lateral gene transfer
µl	Microliter
min	Minute
mM	Millimolar
MOCV	Molluscum Contagiosum virus

MOI	Multiplicity of infection
MPXV	Monkeypox virus
MSDdRP	Multisubunit DNA-dependent RNA polymerase
MV	Mature virion
MYX	Myxoma
MYXA8	vTF7-3 containing the MYXV A8 in place of the native VACV gene
MYXA23	vTF7-3 containing the MYXV A23 ortholog in place of the native VACV gene
MYXV	Myxoma virus
NCLDV	Nucleocytoplasmic large DNA virus
ng	Nanogram
NHP	Non-human primate
nm	Nanometer
nt	Nucleotide
NTPase	Nucleoside triphosphatase
OPXV	Orthopoxvirus
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen-associated molecular pattern
PAS	Polyadenylation site
PBS	Phosphate-buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction

PFU	Plaque-forming unit
PI	Phosphatidylinositol
PKR	Protein kinase R
Pol	Polymerase
PS	Phosphatidylserine
RACK1	Receptor for activated C kinase 1
RAP94	RNA polymerase-associated protein of 94kDa
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNAP II	RNA Pol II
RPO	Viral RNAP subunit
rRNA	Ribosomal RNA
s	Second
SDS	Sodium dodecyl sulfate
SGPV	Salmon Gill poxvirus
SIE	Superinfection exclusion
SINE	Short interspersed element
SIV	Simian immunodeficiency virus
SFV	Shope fibroma virus
SNP	Single nucleotide polymorphism
SPV	Sheeppox virus
TATV	Taterapox virus

TEM	Transmission electron microscopy
TF	Transcription factor
TM	Transmembrane domain
TOPO IIA/B	Topoisomerase IIA/B
TPR	Tetratricopeptide repeat
ts	Temperature-sensitive
TSS	Transcription start site
UDG	Uracil DNA glycosylase
UTR	Untranslated region
UV	Ultraviolet
VACV	Vaccinia virus
VARV	Variola virus
VEGFA	Vascular endothelial growth factor A
VETF	VACV early transcription factor
VITF	VACV intermediate transcription factor
VLTF	VACV late transcription factor
VMAP	Viral membrane assembly protein
vRNAP	Viral RNAP
VTF	Vaccinia termination factor
vTF7-3	WR strain of VACV that encodes the bacteriophage T7 RNA Pol regulated by a VACV early/late promoter
WGS	Whole genome sequencing
WHO	World Health Organization
WR	Western Reserve strain of VACV

WT or wt	Wild-type
WV	Wrapped virion

Chapter 1: Introduction

Poxviruses (POXVs) are a family of large double-stranded DNA (dsDNA) viruses that are able to replicate entirely in the cytoplasm of the cells they infect (7), a feature which distinguishes them from most other DNA viruses that must replicate in the nucleus of cells. The most notorious member of the family is Variola virus (VARV), the causative agent of Smallpox, one of the most devastating diseases in human history. It is estimated that this virus was responsible for no less than 300 million deaths and as many as 500 million during the 20th century (8, 9), while possibly killing off 10% of the world's population over the course of the last millennium (9). Through an aggressive vaccination campaign led by the World Health Organization (WHO) that started in 1967, however, Smallpox was declared eradicated in 1980 (8-15) with the last natural case of the disease occurring in 1977 (16) in Somalia (9, 11). As such, it is still the only human infectious disease to be successfully eradicated (11), and only one of two infectious diseases to have been eradicated, with Rinderpest being the second (17-19). Recent reports, however, have revealed an outbreak of the latter disease in Bulgaria (20), (21), though this outbreak is being caused by an ovine (sheep and goats) rinderpest virus that is closely related to the bovine (cattle) rinderpest virus, which was the one eradicated (21).

Despite the eradication of Smallpox, the need to study POXVs remains. Molluscum Contagiosum virus (MOCV), like VARV, is an obligate human pathogen, though it generally only causes benign lesions in young children. There are also other Smallpox-like diseases that are capable of infecting humans, such as Monkeypox

virus (MPXV) (22, 23), which is responsible for outbreaks in Central Africa. Other POXVs, such as Vaccinia virus (VACV) and Cowpox virus (CPXV), are endemic in parts of Brazil (12) and Europe, respectively, and are also capable of infecting humans via zoonotic transmission, similar to MPXV (22-24). The emergence and spread of these diseases, in combination with the fact that newer generations are not vaccinated against Smallpox and older generations who are protected are dying off, mean that the threat of POXVs as infectious agents is ever-present (8, 22). This threat is enhanced in light of the potential use of Smallpox, and perhaps similar diseases, in bioterrorism attacks (12, 22, 23). There have also been multiple reports of vaccine or VACV escapees infecting animals and establishing enzootic cycles, leading to epizootic infections of horses and other animals (12, 24). Just as alarming is the occurrence of cases in which vaccinated individuals have transmitted the virus to animals (12, 23, 24). As such, learning as much as we can about them will help when POXV outbreaks occur.

POXVs have been studied for other reasons though. For decades they have been used as expression vectors, which has provided a powerful tool for immunologists and biochemists alike (7), and as gene delivery systems for gene therapy, due to the relative ease with regard to formation and isolation of recombinant viruses, their relatively high gene expression, their wide host ranges (7), and their large genomes and ability to incorporate large fragments of DNA (7, 25-27). They have also been considered and/or developed for use as both animal and human vaccines and oncolytic viruses in cancer therapy (7, 11, 23, 27-33). POXV research has also contributed significantly to basic science, providing insights into virus-host

interactions and general molecular biology (7, 34, 35). VACV was the first animal virus to be seen through a microscope, grown in tissue culture, accurately titered, physically purified, and chemically analyzed. Discovery of RNA-synthetic activity and virally-encoded proteins which affect cell growth and modulate host immune defense mechanisms led to discoveries of similar activities in RNA viruses and new insights into virus-host relationships, respectively (7).

This work focuses on the interesting gene expression system of poxviruses, which is separated into three distinct phases: early, intermediate, and late. Each phase has its own stage-specific transcription factors, along with stage-specific promoter sequences, which allow for expression of genes at certain times (7). Currently, much more information is known regarding early transcription and its regulation compared to intermediate and late transcription. Chapter 2 will expand on this disparity in knowledge as well as the differences between early and intermediate/late transcriptional regulation, in addition to how POXVs adapt to selective pressures. Chapter 3 will detail experiments used to determine if orthologs of the VACV intermediate transcription factors are able to complement viruses deficient in these genes. Promoter sequences are quite similar between POXVs (27, 36-38), so we wanted to see if the transcription factors were equally interchangeable. Once that was determined, recombinant viruses with orthologs in place of the native VACV genes were generated and blindly passaged to see if mutations would arise that would enable these recombinants to replicate better. Whole genome sequencing (WGS) discovered the presence of five distinct mutations present in the two largest subunits of the vRNAP and two additional mutations near the translation initiation

site of the Myxoma A8 ortholog. Experiments were then carried out which confirmed that these mutations were indeed responsible for the increases in fitness of the recombinant viruses. The experiments in Chapter 3 were recently published. My contributions to this work included generating the deletion viruses and plasmids containing the orthologs, performing the initial Western blots and Luciferase assays, generating the hybrid viruses, carrying out the blind passaging, isolating the P10 clones, preparing and sending the viral DNA for WGS, helping with the Amplicon Sequencing, and carrying out the viral yield assay. Appendix A will explore the preliminary experiments carried out to characterize these mutations by attempting to determine the effects of these mutations on the virus.

Chapter 2: Literature Review

2.1 The Poxviridae

2.1.1 Classification

The family *Poxviridae* is distinctive thanks to their cytoplasmic sites of replication, large complex virion, and single, linear dsDNA genome ranging from 130 to 300 kilobase pairs (kbp) in length, with a hairpin loop at each end (7). They are placed in Group I: dsDNA of the Baltimore classification system (29). POXVs are distantly related to other large DNA viruses, in particular members of the following virus families: *Asfaviridae*, *Iridoviridae*, *Phycodnaviridae* (7), and *Ascoviridae*, all of which were discovered prior to 1983 (4). All of these viruses are also members of the superclade, proposed order *Megavirales* (2), of viruses known as nucleocytoplasmic large DNA viruses (NCLDVs), all of which infect eukaryotes (4, 39). Since 2003, more viruses and virus families have been discovered whose characteristics led them to be categorized as NCLDVs. They are *Mimiviridae*, *Marseilleviridae*, *Megaviridae*, *Pandoraviridae*, *Pithoviridae*, and *Faustovirus* (4).

POXVs themselves are divided into two subfamilies: *Chordopoxvirinae*, which infect vertebrates, and *Entomopoxvirinae*, which infect insects. Entomopoxviruses are divided into three genera, based on the insect host of isolation: *Alphaentomopoxvirus (7 species)*, *Betaentomopoxvirus (13 species)*, and *Gammaentomopoxvirus (6 species)* (7, 40). There are also two unassigned species (40). Chordopoxviruses (ChPVs) (32) were previously divided into eight genera: *Orthopoxvirus (9 species)*, *Parapoxvirus (4 species)*, *Avipoxvirus (10 species)*, *Capripoxvirus (3 species)*, *Leporipoxvirus (4 species)*, *Suipoxvirus (1 species)*,

Molluscipoxvirus (1 species), and *Yatapoxvirus (2 species)* (7, 40). Since then, two more genera have been added: *Cervidopoxvirus (1 species)*, in addition to one unassigned species (40), Crocodilepox virus (41, 42). Viruses within a genus are closely related genetically, and have similar host ranges and morphologies (7). All ChPVs have their genes similarly arranged, have interchangeable promoters, and have conserved RNAP and transcription factors (TFs) (27). Members of the Parapoxvirus genus have a broad host range and infect animals in the Laurasiatheria (placental mammals) superorder, including humans occasionally. Viruses in the Avipoxvirus genus on the other hand are only able to infect birds, although abortive infections have been observed in other animals. Member viruses of the Capripoxvirus genus infect ruminants, which include cattle, sheep, and goats. Species of the Leporipoxvirus, Suipoxvirus, Molluscipoxvirus, and Cervidpoxvirus genera are each only capable of infecting one type of organism or species: rabbit or leporid species, swine, humans, and species of deer, respectively. Finally, Yatapoxvirus genus viruses infect primate species (42). Prior to the advent of genome sequencing, it was often theorized that the association of poxviruses with their vertebrate hosts was quite ancient and thus host and parasite had co-evolved throughout the history of amniotes. Phylogenetic analysis in general, however, does not support this analysis and instead strongly hints that transfers from one host species to another have been a recurrent feature throughout the evolution of the *Chordopoxvirinae* and that, overall, host and POXV lineages have not co-evolved (43).

2.1.2 Orthopoxviruses

Of the ChPVs, the Orthopoxviruses (OPXVs) have been the most extensively studied, with VACV serving as the prototypical member of the genus (7) and, as a result, the best characterized virus in the NCLDV superclade (4). Viruses in this genus are immunologically cross-reactive and cross-protective, so infection with one member confers immunity against another member of this genus (8). As a genus, OPXVs are believed to have diverged from the other POXV genera about 130,000 years ago (44). There are 10 virus species which make up the OPXV genus: Camelpox (CMLV), CPXV, Ectromelia (ECTV), MPXV, Raccoonpox, Skunkpox, Taterapox (TATV), VACV, VARV, and Volepox. The CPXV genome contains all the genes found in other members of the OPXV, suggesting that it is the least diverged from the ancestral member of this particular genus (1, 7, 44, 45). The working hypothesis is that the subsequent evolution of the OPXVs has been a gradual restriction of host range via mutations of genes and loss of genes/shortening of the genome (1, 45, 46). As a result of having the largest genome, CPXV has the widest host range (1, 13, 44, 45). Additionally, recent analyses suggest that multiple species of OPXV have been grouped under the CPXV species (40, 44), which can be separated into anywhere from three to five distinct phylogenetic clades (13, 24). Many strains began evolving independently from a common ancestor about 10,000 years ago (44). The diversity among these isolates is so great it has been proposed that they should be assigned to different OPXV species (1, 44).

Genes common to VACV, VARV, CPXV, ECTV, and CMLV all have greater than 90% sequence identity (7). Conversely, OPXV native to the Americas, such as

Raccoonpox, Skunkpox, and Volepox, show a much greater genetic divergence (40), with their progenitor virus diverging about 50,000 years ago after the New World was colonized (44). Viruses in the OPXV genus have a broad host range and infect a wide variety of mammals, including humans. Species of CPXV infect many mammals, such as cats (13), including humans (unvaccinated ones in Europe (13)), and even zoo animals (13), but their natural hosts seem to be rodents. MPXV species also have a wide host range, infecting non-human primates (NHPs), humans, and other large animals, in addition to a large variety or rodents. As with CPXV, rodents may be the natural hosts for these viruses. The species origins and natural host(s) of the VACV species are currently unknown (14, 24, 42), though the natural source is assumed to be Horsepox virus (HSPV), which can also infect rodents (1). In laboratory settings, VACV species infect a wide range of animals. This wide host range is highlighted by the occasional outbreaks among bovine populations in South America (12, 42) and the ability of the virus to readily pass from animal-to-human and vice versa, as mentioned above (12, 23, 24). CMLV virus is only known to infect camels, while TATV virus infects gerbils. ECTV infects mice and moles (23, 42). VARV, like MOCV, is a human-specific pathogen (42).

2.1.3 Virion Chemical Composition

The VACV mature virion (MV) is the basic and most abundant infectious form of the virus. Also known as the intracellular mature virion (IMV) (47, 48), it has a mass of 9.5 femtograms (fg). The main components are proteins, lipids and DNA, which represent 90%, 5%, and 3.2% of the dry weight, respectively. In contrast, the prototypical member of the Avipoxvirus genus, Fowlpox (FWPV), has

about one-third of its virion weight in the form of lipids. The lipids in the VACV MV are primarily cholesterol and phospholipids (40, 49, 50), while FWPV also contains squalene and cholesterol esters. Spermine and spermidine, along with trace amounts of RNA have also been found in the VACV MV (7), which may lend credence to the results published in a recent paper (51).

2.1.4 Virion Morphology

POXV virions are quite large for animal viruses. They are so large that they are just discernible using a light microscope. To see ultrastructure, however, electron microscopy (EM) is required. This tool, along with a technique known as Cryo-electron tomography, have provided many details about the organization and composition of the POXV virion. The VACV MV possesses a barrel shape, dimensions of approximately 360x270x250 nanometers (nm), and an outer layer whose density and thickness (5-6 nm) seem consistent with one lipid bilayer membrane. The complex internal structure consists of a dumbbell-shaped core and aggregates of heterogeneous material, known as lateral bodies, occupying the space between the concavities and the outer membrane. The core wall, which has an overall thickness of 18-19 nm, consists of two layers: an inner layer and an outer layer. The inner layer, which appears continuous, except for a small number of channels, has a diameter and density consistent with a lipid membrane. The outer layer on the other hand has a palisade structure, made up of T-shaped spikes (8 nm long and 5 nm wide) which are anchored in the putative lower membrane. Freeze fracture and deep etch EM have been used to confirm both the outer membrane of the virion and the palisade structure of the core wall, but do not support the existence of an internal core

membrane. Cryo-electron tomography, while a powerful technique, has its limits and is not yet able to resolve the ultrastructure of the core. It has been able to glean, however, that the core has two distinct phases: the core wall itself and a denser layer underneath with a fiber-like morphology, suggesting nucleoprotein. Cylindrical elements, which may take on an S-shape or more complex, flowerlike structures, have been visualized using other methods within POXV cores (7, 40). These structures are presumed to represent nucleoprotein (7), though no viral proteins have yet been localized to structural features of virions (40). A recent publication has found that viral transcripts are incorporated into virions, at least in the case of CPXV (51).

2.1.5 Polypeptide Components of the Virion

POXV virions contain many polypeptides. From MVs alone about 30 can be visualized using one-dimensional polyacrylamide gel electrophoresis (PAGE), after being purified from infected cells and disrupted with sodium dodecyl sulfate (SDS) and a reducing agent (7). About 80 can be detected using a two-dimensional gel or mass spectrometry, while 30 have been characterized as being near the exterior of the purified MVs using one or a combination of techniques, e.g. surface-specific labelling, sensitivity to proteases, extraction with nonionic detergents, and reactivity with neutralizing antibodies (40). These surface proteins can be categorized, structurally, as those with or without transmembrane domains (TMs) (7), and functionally, as those required for attachment, entry, disulfide bond formation, morphogenesis, and virulence (40). The source of the viral membrane has since been found to be the endoplasmic reticulum (ER) (52-55).

Treating MVs with a nonionic detergent and a reducing agent leads to the release of viral cores. Almost 50 polypeptides have been identified in the cores, none of which can be found in the membrane fraction. Enzymes make up about 30 of these, and of these 30, at least half are directly involved in the biosynthesis of early mRNA. The remaining non-enzymatic proteins could be involved in morphogenesis and structure, with A17, L4, A3, and A10 being the most abundant (7).

Meanwhile, the VACV enveloped virion (EV), which is its other infectious form, contains carbohydrates as a result of its glycoproteins (7). The EV can also be referred to as the cell-associated enveloped virion (CEV) or extracellular enveloped virion (EEV) (47, 48). The EV possesses an additional membrane, which contains several glycosylated proteins unique to this extra membrane: A33, A34, A56, B5, and K2, and F13, which is a non-glycosylated putative phospholipase (7). EVs are also reportedly enriched with sphingomyelin and phosphatidylserine (PS) with lower amounts of phosphatidylinositol (PI), relative to the MV (47). This extra membrane results in EVs having a lower buoyant density than MVs (40). Of the six EV membrane proteins, only two, A34 and F13, have been found in all ChPVs. While EVs are generally regarded as simply being MVs with an additional membrane, this definition is not technically accurate. Evidence suggests that MVs possess at least two surface proteins that EVs do not, A25 and A26. The existence of EVs was first suggested by vaccine-related studies, where it was noticed that antibodies against inactivated MVs failed to protect rodents as well as antibodies to live virus or inactivated EVs in the context of an OPXV infection. In addition, there were

differences between the different types of sera in their abilities to neutralize virus (40).

2.1.6 Genome Organization

All POXVs contain a linear, dsDNA genome. These genomes can range in size from 133-134 kbp in Parapoxviruses and Yatapoxvirus (42) to 300 kbp in some Avipoxviruses (7), including almost 360 kbp in Canarypox virus , and even larger than that in some Entomopoxviruses (370-380 kbp) (7), and encode usually more than 150 genes (56). This coding potential can range, though, from 133 genes in Parapoxviruses and Yatapoxviruses to 328 genes in Canarypox virus (32, 42). The average size of a POXV-encoded protein is about 30 kiloDaltons (kDa or kD), similar to eukaryotes. Furthermore, as these viruses replicate in the cytoplasm and thus do not undergo mRNA splicing, the number of viral genes can be estimated to be 1 per 1 kb of genome (56). The genome itself is not infectious, as virally encoded enzymes and other factors are required for expression in the cytoplasm. Complete genome sequencing has been done for two Entomopoxviruses and at least one member of each ChPV genus (7). Out of all the POXV genomes that have been fully sequenced, 90 genes are conserved in all ChPVs, which represent about half of the genes encoded by the ChPV subfamily (48, 50), with about half of these, 41, also present in Entomopoxviruses (48, 56, 57).

The organization of POXV genomes appears to follow some general rules: genes for the most part do not overlap, tend to occur in blocks pointing towards the nearer end of the genome, and ones essential for replication, which are highly conserved, are located in the middle of the genome, while those that are needed for

host interactions, which are variable and species-specific, are located on the ends of the genome (7) (**Fig. 2.1**). A significant proportion of the variable gene products engage host defense mechanisms (50), while the genes in the central region are vital for cell entry, DNA replication, gene expression/transcription, intramolecular disulfide bond formation, and virion assembly (48, 50). The arrangement of the central genes is similar across all ChPVs (7). This core region of the genome makes up about 75% of the complete sequence (42). One study found that both gene order and gene spacing are well conserved within the ChPV genomes, with the FWPV genome being a notable exception, as it has undergone several rearrangements (58). Another group found that this exception applied to other Avipoxvirus genomes as well, as large genome inversions were found (42). The former study also found that, of the genomes it analyzed, there is seemingly no conservation of gene order between Entomopox genomes or between the Entomopox and ChPV genomes (58).

All POXV genomes contain inverted terminal repeats (ITRs), which consist of identical yet oppositely-oriented sequences of the genome. These regions include: an AT-rich, incompletely base-paired, hairpin loop that connects the two DNA strands and results in their linear genome; a highly-conserved region of almost 100 base pairs (bp) that contains sequences required for the resolution of concatemers formed during genome replication; variable-length sets of short, tandemly repeated sequences (7, 59-62); and up to several open reading frames (ORFs). These ITRs end up being variable in length as a result of a combination of deletions, repetitions, and transpositions (7). The ITRs vary in size between species, with some being sufficiently large as to contain the coding region for multiple genes, meaning that

genes within the ITRs are present as diploids in the viral genome. For instance, within the OPXV genus, the ITRs of VARV are about 200-500 bp and contain no genes, while strains of VACV have ITRs that are almost 12 kbp in length and contain six diploid genes (42).

2.1.7 Genome Nomenclature

The practice of naming VACV genes or ORFs was used prior to the entire genome being sequenced but has since been used for the complete sequence of the Copenhagen strain of VACV (VACV CP). The convention used called for the naming of fragments of the VACV genome generated after digestion with the HindIII restriction endonuclease. Names were assigned based on the size of the fragment, with the largest fragment referred to as A, the second largest referred to as B, and so on. Within each fragment, each ORF was given a number, from left to right, followed by an L or R, depending on which direction the ORF was transcribed (7, 48). The exception to this rule was the C fragment, which was numbered from right to left to avoid starting in the highly variable left end of the genome (7). The polypeptides that are encoded in the genome share names with their respective ORFs, but the L/R designation is dropped (7, 48). Regarding the subsequent complete POXV genome sequences, the ORFs were successfully numbered from one end to the other. Despite this, the old letter designations are still used as common names when referring to homologs to avoid confusion and to provide some continuity in the POXV literature. The ORF number of the Western Reserve (WR) strain of VACV is widely used as this strain has been used for the vast majority of biochemical and genetic studies (7). Restriction enzyme mapping of POXVs is also useful for

generating physical maps of these viruses (63), which can aid in identification and classification of POXV isolates as more closely related POXVs tend to give similar banding patterns when digested with the same enzyme, while more distantly related POXVs yield dissimilar banding patterns when cut with the same enzyme (64).

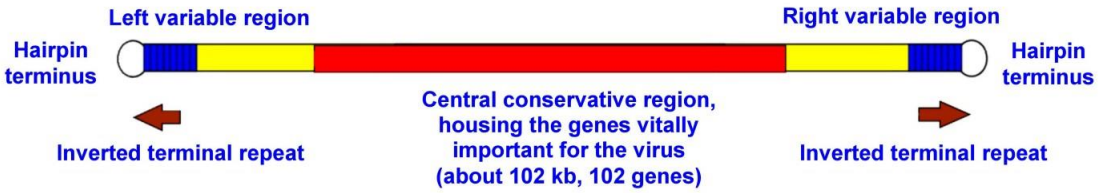

Figure 2.1. Scheme of OPXV genome (not to scale). Figure 1 from Babkin, I.V. and Babkina, I.N. The Origin of Variola Virus. Viruses, 2015. 7: 1100-1112. (1). Reprinted with permission under the Creative Commons Attribution License.

2.2 Virus Lifecycle

2.2.1 Attachment and Entry into Cells

There are many mysteries surrounding POXV attachment and entry into cells, compounded by the fact that there are two forms of infectious particles, the MV and EV. The surface proteins between the two are different and evidence supports the idea that they have distinct binding sites. This includes relative differences in binding efficiencies of the two virus types to specific cell lines, differing effects of changes in binding after cell surface digestion with proteases, and the effect of a monoclonal antibody against the cell surface on the ability of MVs to bind to it (7). Even though the MV is the more abundant form of the virus, the EV is the form more adapted for cell-to-cell spread (40).

2.2.2 MV Entry

Evidence suggests MVs utilize one of two pathways for entry. One involves fusing with the plasma membrane, which demonstrates a pH-neutral mode of entry. The second involves endosomal entry, indicating a low-pH method of entry (7, 48). Endosomal acidification enables entry of the viral core into the cytoplasm, following actin-dependent micropinocytosis or fluid uptake of the large viral particles. Myxoma virus (MYXV) virus also enters cells via a low-pH endocytic pathway (40). Which method is used may depend on the strain of virus and cell type being studied (7) as there is considerable variation with regard to preference for alternative entry pathways when it comes to strains of VACV. This preference may be dependent on the A25 and A26 proteins. The versatility of VACV to utilize more than one entry

mechanism could explain their ability to infect most cell types. For example, VACV infects primary cultures of well-differentiated human airway epithelial cells via the basolateral surface but egress via the apical surface (40). VACV can even enter Drosophila S2 cells and other insect cells, although replication is abortive (48) and does not proceed past DNA synthesis (65). Dual entry pathways have also been observed in herpesviruses (7). Endosomal entry is thought to be more advantageous as it allows for passage through the dense cell cortex. Evidence suggests that this is the preferred pathway for the WR strain of VACV (48) and of OPXVs that have recently been isolated in nature (50). The A25 and A26 proteins reportedly serve as fusion suppressors for MVs, dictating the entry pathways utilized by specific strains, i.e., MVs lacking these two proteins enter through the plasma membrane, while those containing functional versions of these proteins do not readily fuse with the plasma membrane and thus enter via the endocytic pathway. The route of viral entry is also cell-type dependent. For example, VACV infection of Drosophila S2 cells happens exclusively via the endocytic route (48). Using fluorescence microscopy, MVs have been observed trafficking to early endosomes and recycling endosome compartments prior to fusion of the membrane, following endocytosis (50). Despite their biological significance, very few studies on VACV entry have used EVs and so most of these studies have focused on MVs (48, 50). The reasoning is two-fold. From a practical point of view, the fusion proteins are only located on the MV membrane (50). From a technical aspect, a high proportion of EVs released from cells contain partially disrupted outer membranes, making results obtained with these particles difficult to interpret (48).

2.2.3 Attachment Proteins

Many enveloped viruses possess one or two proteins which serve dual roles in attaching to and fusing with cellular membranes. VACV is unique in that it encodes at least four attachment proteins and 11 entry proteins (40). The four attachment proteins are A26, A27, D8, and H3. The initial attachment appears to be facilitated by binding to glycosaminoglycans. D8 can bind chondroitin sulfate, A27 and H3 bind to heparan sulfate (7), and A26 binds laminin. Of these attachment proteins, only two have a TM domain, D8 and H3. A26 and A27 interact with each other, the latter of which is anchored by the A17 TM protein (40), an important component of the MV membrane (48) that will be discussed later. Individually, none of these proteins are essential (7, 48), although deletion of A27 and H3 severely reduces VACV infectivity (48, 50). The crystal structures of both A27 and H3 have been solved, with the former in the form of a trimer and the latter in the form of a dimer (50). All four attachment proteins are multifunctional (48), having roles unrelated to attachment (50), but only H3 is highly conserved among POXVs. H3 is involved in MV assembly, while A27 is required for wrapped virion (WV) formation. A26 mediates MV incorporation into structures known as A-Type cytoplasmic inclusion bodies (ATIs) of some OPXVs, such as CPXV and ECTV. The VACV A25 protein is a truncated form of the ATI protein, but is reported to be a fusion suppressor. A soluble, truncated, recombinant form of the L1 protein, which will be described in more detail later, can attach to cells lacking glycosaminoglycans. This results in the blocking of VACV entry, suggesting it is a receptor-binding protein. The putative cellular protein responsible for this interaction, though, has yet to be found (48).

Interestingly, heparin, which has an inhibitory effect, appears to exert a greater competitive effect on strains which preferentially enter via a pH-neutral mechanism (40). While the viral attachment proteins are known for the MV, those associated with the EV have not yet been identified (48). EM images have shown dissociation of the EV membrane at the cell surface though, supporting the idea that it is not fusogenic. Rupture of this membrane requires the EV membrane glycoproteins A34 and B5, along with interactions with glycosaminoglycans on the cell surface following virus attachment (50).

2.2.4 Entry Proteins

An entry-fusion complex (EFC), a group of molecules essential for viral entry, was identified, and it consists of at least 11 proteins: A16, A21, A28, G3, G9, H2, J5, L5 (40), F9, L1 (66), and O3 (48, 67). L1 and F9 are associated with the EFC, while the rest are all integral components of the complex (48). All 11 are necessary for entry (48) but not for assembly of virions or attachment to cells (7, 48), as assembly of normal-looking virions still occurs if these proteins are repressed (40, 48). These viruses, however, are unable to infect cells. More evidence of their importance is the fact that all 11 proteins have been found in all POXVs analyzed to date, suggesting a common entry mechanism (40). These proteins are all made after DNA replication (48), range in size from 4 to 43 kDa, are not glycosylated, and do not resemble the type 1, 2 (40, 48), or 3 (50) fusion proteins found in other viruses (40, 48, 50). The lack of glycosylation and cleavable signal peptides raise questions regarding how these proteins are trafficked to the MV membrane (50). All, with the exceptions of G3 and O3, have conserved intramolecular disulfide bonds that are formed by the

POXV-encoded cytoplasmic redox system, which has also been found in all POXVs to date (40, 48). As no other viral proteins have yet been identified as substrates of the POXV redox system, the current thinking is that this system co-developed with the EFC proteins (48). Three of the proteins appear to have been derived from a common gene early in POXV evolution: A16, G9, and J5 (40, 48). This seems to be the result of a gene duplication, which then diverged over the course of POXV evolution (48). These three proteins have low sequence identities, despite their evolutionary relationship (50). In that same vein, L1 and F9 are structurally related. Despite this, each paralog is encoded by all POXVs and each one is individually required for entry (48). All 11 proteins possess a TM domain (66), with five possessing it in their N-terminal regions (A21, A28, G3, H2, and O3), while the remainder (A16, F9, G9, J5, L1, and L5) have it in their C-terminal regions (48). The proteins that make up this complex are required for the membrane fusion step of entry of the MV (66), cell-to-cell spread of the EV, and low-pH-induced syncytia formation, all of which support the idea that only the MV membrane is fusogenic and that fusion and entry are coupled together (7).

While some interacting partners have been identified, neither the stoichiometry nor the structure of the EFC has been solved (40). This complex fails to come together if viral membrane formation is inhibited, which prevents its synthesis in heterologous systems, thus hindering its physical characterization. The combined mass of the EFC and the associated proteins would be 232 kDa, operating under the assumption that each protein is only present in a single copy. The EFC is also destabilized when any one of the nine integral proteins is repressed, hinting at it

being held together by multiple subunit interactions, some of which have been solved. These interactions are: A28-H2, A16-G9, and G3-L5. The A16-G9 complex has an additional function, as it can bind to the A56-K2 heterodimer of fusion regulatory proteins as well as A26 (48). O3 is the smallest protein encoded by VACV, at just 35 amino acids (aa) long. Orthologs from other POXVs, ranging in in size from 29 – 48 aa in length, can complement an O3 deletion mutant, despite their low degree of amino acid identity. This is thanks to its N-terminal TM domain, which is essential but sufficient for it to associate with other EFC proteins (48, 67). Mutagenesis experiments with H2 identified a highly-conserved region (a conserved LGYSG sequence between two cysteines (50)) that is crucial for its interaction with A28. L1 is myristoylated at a Glycine residue in its N-terminal region and has three intramolecular disulfide bonds. Mutating this Glycine prevents the complementation of VACV infectivity, changes the intracellular localization of L1, and reduces the formation of the intramolecular disulfide bonds. Despite all that, the mutated protein will still associate with the EFC and MVs. The crystal structure of this protein has also been solved, revealing a fold comprised of a bundle of α-helices packed against a pair of two-stranded β-sheets, as well as a large hydrophobic cavity which would seem to accommodate the N-terminal myristate moiety. Mutations within said cavity have no effect on myristoylation but do inhibit infectivity, leading to a model involving a "myristate switch", in which the acyl chain is released from the cavity during entry. Interestingly, both L1 and A28 are targets of neutralizing antibodies, meaning that they are both at least partly exposed on the surface of the MV. The immunogenicity of A28 is actually enhanced when it is associated with H2 (48).

Additional proteins, including A27 and I2, have been implicated in entry (7). Specifically, A27, along with A17, have been implicated in membrane fusion, but no genetic evidence supports their requirement for entry (40, 48). The heterologous expression of A17 reportedly causes the fusion of transfected cells, suggesting a similar role in viral entry. Conditional lethal A17 mutants have a block in viral membrane formation, where fusion may play a role, making it hard to determine an additional entry function (48). With regard to I2, it was found that virions lacking the protein are unable to enter cells, thus greatly reducing infectivity (66). This Hyun et al. (66) study set out to investigate the initial report from another group regarding this defect in viral entry by determining which step of replication or entry was affected by the absence of I2. I2 itself is a small protein, at only 72 aa long and with a mass of 8.4 kDa. It possesses a C-terminal TM domain, does not contain any conserved intramolecular disulfide bonds (48), is made after VACV DNA replication, is associated with purified MVs, and is conserved in all ChPVs. What they found was that, while an I2 deletion mutant was unable to enter cells, the primary block was, unexpectedly, in virion morphogenesis, specifically after the formation of IVs, a process which will be discussed in more detail later. Virions lacking I2 were deficient in EFC proteins, which likely prevented the entry of the aberrant particles into cells. Given the findings, they concluded I2 is not an entry protein per se but has an indirect effect on viral entry.

2.2.5 EV Entry

The entry of EV particles requires at least one additional step, as the outer membrane must be shed for fusion to occur since the entry-fusion proteins are found

only on the MV membrane. This additional membrane contains several proteins unique to it, most of which are glycosylated, A33, A34, A56, B5, and K2, and one which is not, F13, which is a non-glycosylated putative phospholipase. EVs appear to be able to shed their outer membranes at both low and neutral pHs. The A34 and B5 glycoproteins, plus polyanionic molecules, are required for this process at neutral pH. B5 has also been implicated in cell entry, thanks to neutralizing antibody studies. Mutations in either member of the A56 hemagglutinin-K2 glycoprotein heterodimer leads to fusion of adjacent infected cells without the need for a low pH trigger. EVs are largely responsible for the efficient cell-to-cell spread of VACV by adhering to the cell surface at the ends of finger-like projections that are formed by actin polymerization. Actin has also been potentially implicated in MV entry via an actin-mediated internalization mechanism, perhaps involving macropinocytosis. For EVs, deletion of the A34R, A33R, or A36R genes prevents the formation of actin tails and reduces virus spread efficiency and EVs lacking A34 possess low infectivity. Deficiencies in actin tail formation can be partly overcome using mutants or strains that produce more EVs. One such example is the IHD strain of VACV, which seems likely to have arisen from a spontaneous mutation in the A34R gene, releases large amounts of EV particles and produces comet-like satellite plaques (7). Other viruses which possess mutations in A33, A34, or A36 are able to overcome a deficiency in actin tail formation via enhanced EV production (40).

2.2.6 Signaling Receptors

Studies have indicated that the interaction between the MV and the cell surface triggers a signaling cascade which leads to membrane rearrangements along

with the formation of actin- and ezrin-containing protrusions that envelop the virus (7). MVs are engulfed by the cell via clathrin- and caveolin-independent micropinocytosis or fluid phase endocytosis, both of which are dependent upon actin dynamics and cell signaling (48). Additional experimental evidence suggests that both signaling and actin rearrangements are required for the initial fusion of the viral and cellular membranes (40). Inhibitors of actin dynamics do not significantly affect binding but do impair membrane fusion. It has been proposed that actin remodeling facilitates fusion by forcing membranes together and enlarging pores in certain situations, such as viral protein-induced cell-cell fusion and virus entry. In fact, actin dynamics may be required for both hemifusion and pore formation. The requirement for cell signaling may have been the impetus behind earlier proposals putting forth chemokine receptors as the receptors for POXV entry. It has been determined, though, that the cellular proteins, VPEF (48), integrin B1 (50), and CD98 (48, 50), participate in the fluid phase uptake of MVs (48, 50). These proteins are associated with lipid rafts (48). In contrast to this, however, is EV entry, which does not appear to be signaling dependent (47), which may occur using either of the above routes used by MVs (48, 50).

2.2.7 Cellular Lipid Composition

The lipid composition of the cell membrane is important for entry, as virus penetration is inhibited by the depletion of cholesterol (7, 47). Cellular cholesterol inhibition does not prevent binding but it does significantly reduce viral core entry, possibly suggesting a role for lipid rafts (47, 48). MVs have reportedly been found to associate with cholesterol-rich regions of the plasma membrane (50). Inhibitors of

endosomal acidification and membrane blebbing resulted in the same phenotype (48). Specifically, this block occurred during the initial stages of virus-cell membrane fusion. These experiments also supported the hypothesis of a two-step entry model for VACV with a hemifusion intermediate (49). In this model of entry, the outer leaflets of the apposing membranes form a hemifusion intermediate, followed by merging of the inner leaflets, thus forming the fusion pore. As such, lipid mixing and content mixing occur in a sequential order (48, 49). The site of hemifusion, although yet to be directly determined, is thought to be at the plasma membrane (50). The individual roles of each EFC member were investigated using conditional lethal inducible mutants. All were able to attach to cells but unable to carry out even the initial step in the fusion process, with the exceptions of the A28, L1, and L5 mutants. For these three, lipid mixing occurred, but no subsequent steps took place. These data further supported the two-step model of viral entry (48, 49).

VACV infectivity can be enhanced in combination with PS and an apoptotic mimicry model with a role for a specific PS receptor has been put forth, based on reconstitution of delipidated virus (40, 49, 50). In this model, virion-associated PS flags virions as apoptotic debris for cell uptake via macropinocytosis (49, 50). The stereoisomer of PS and other phospholipids not known to signal apoptotic uptake are also capable of reconstituting infectivity however (40, 49, 50). This means that either the putative receptor has broad specificity or that the phospholipids play a different role in enhancing infectivity (40, 68). Regardless, the viral membrane lipid composition, specifically the presence of anionic phospholipids, is important for the entry of VACV MVs into cells (47, 50). Entry of the core is the last step in

membrane fusion. Cores can be detected in the cytoplasm of infected cells in as little as 10 minutes (mins), when the medium is acidified and entry occurs via the plasma membrane. At neutral pH, however, it takes 20-30 mins to detect cores being liberated from the endosome. These determinations were done using fluorescence and electron microscopy. Using a luciferase assay-expressing recombinant VACV, there is a delay of about 20 mins in detection of cores by EM versus detection of an appreciable amount of luciferase activity. It was determined that at least 10 of the EFC proteins are required for core entry (50).

Specific protein receptors required for viral entry have not been identified yet, though some have been proposed, such as the epidermal growth factor receptor and chemokine receptors (7, 48). Neither of these theories has been supported by experimental results however. Slow progress on this aspect of the POXV lifecycle is due, at least in part, to the fact that no cell line refractory to POXV entry has yet been identified. Once such a cell line was found, a genetic screen could be performed to identify a receptor, as has been done with numerous other viruses (7). It should be noted, however, that for VACV and most other OPXVs, it is believed that cellular tropism is not regulated at the level of binding and entry but rather by intracellular events (69), a theory which has been supported by experimental findings (48, 65, 70).

2.2.8 Syncytia Formation

Syncytia formation is a phenomenon in which cells fuse to one another. It is dependent upon the formation and externalization of virions and on components of the EFC, implying that it has features related to viral entry. This process is also referred to as fusion from within. Infections with only specific VACV mutants can

cause this, namely A56 and K2. These two proteins form a fusion suppressor complex on the plasma and EV membranes and interact with the A16-G9 subunits of the EFC, preventing spontaneous activation of the fusion apparatus by viral progeny. Both A56 and K2 are required for this function, as uninfected cells expressing both, but not either alone, are resistant to syncytia formation when mixed with A56R deletion mutant-infected cells. Syncytia formation can also occur if cells infected with wild-type VACV are briefly exposed to low pH. A similar phenotype is observed with A56R and K2L mutants at neutral pH. This process is also referred to as fusion from within, even though it too is dependent upon virions being present on the cell surface. Fusion from without, on the other hand, can occur when cells infected by a high multiplicity of MVs are exposed to low pH. The A26 protein also reportedly binds to the A16-G9 subunit complex of the EFC. This association is weakened in low pH and A26 deletion mutants are capable of inducing fusion from without at a neutral pH (48, 50).

2.2.9 Superinfection Inhibition

Superinfection exclusion (SIE) is a phenomenon which prevents EVs from infecting cells which express the A33 and A36 proteins, both of which are made by early genes (48), on the surface of their membranes (40, 50). This repulsion mechanism likely enhances the spread of EVs to uninfected cells. Another mechanism POXVs use to prevent superinfection, specific to MVs, is the interaction between the A56-K2 heterodimer and the A16-G9 subunit complex of the EFC, an interaction which also serves to prevent syncytia formation, as mentioned above (40, 48, 50). A56 and K2 expression by infected cells reduces entry of superinfecting

virus at late times, seemingly in addition to the already reduced superinfection exclusion. These two proteins alone are sufficient for superinfection exclusion, as uninfected cells stably expressing both are resistant to infection (48). Thus, this viral protein complex has roles as a fusion suppressor for both viral entry and syncytia formation (50). Yet another way POXVs prevent superinfection is at the membrane lipid-mixing, or hemifusion, step of viral entry. This mechanism requires the primary virus to undergo early mRNA and protein synthesis, but not DNA synthesis. Thus, the secondary, or superinfecting, virus binds to the cell but is unable to release its core into the cytoplasm and initiate early gene expression. This mechanism differs from the A56/K2-mediated block, which takes place after membrane fusion occurs (40, 49, 50). As such, it is necessary to distinguish between SIE of the secondary virus and superinfection resistance induced by the primary virus, which act on different steps of viral entry and via distinct mechanisms (40, 71). Regardless of which mechanism is used, SIE operates between virus adsorption and early gene expression, nearing completion by 6 hours post-infection (hpi) (48).

2.3 Genome Replication

POXVs, as mentioned before, replicate in the cytoplasm of infected cells, and are even able to replicate in enucleated cells. The specific sites or foci of replication are referred to as viral factories. Each infectious particle is capable of generating its own viral factory (7), though factories are capable of merging over the course of an infection (40). DNA replication begins 1-2 hpi and generates about 10,000 copies of the VACV genome, about half of which end up being packaged into progeny virions (7). The timing of DNA replication varies depending on the POXV being studied and

can be affected by the multiplicity of infection (MOI) and cell type (40). One recent study (72) found that, in a confluent A549 cell monolayer, about 500 VACV genomes per cell had accumulated by 6 hpi. This same paper used digital droplet PCR (ddPCR) to determine the rate of replication. The rate of synthesis was calculated to be 86 bp/second (s) from 3.5 to 4 hours (h) and 130 bp/s from 4 to 5 h, which was the maximal rate observed. This rate was several times faster than what has been calculated for herpes simplex virus (HSV) DNA replication. Using these numbers, it was estimated that it would take between 25 and 35 mins to replicate a full-length VACV genome. These calculations were done under the assumption that there is only one replicon per genome. If multiple origins of replication existed, however, the rate of synthesis would be lower (72). There is evidence which suggests replicating DNA may be associated with ER membranes, (40) as the ER is actively recruited as the viral factory expands due to macromolecular synthesis (72). Senkevich et al. (72), however, did not find evidence that VACV DNA synthesis takes place in close association with recognizable ER membranes. They proposed that the ER surrounding the factories is passively pushed out from the interior, in addition to other cellular organelles, as the factory expands.

2.3.1 Uncoating and DNA Synthesis

While it is obvious that uncoating of the viral core is a necessary event prior to DNA replication, more and more evidence has now shown that uncoating and DNA synthesis are deeply intertwined, with a significant role being played by the ubiquitin-proteasome system. The interconnectedness and complexity of these two events were highlighted by the findings of a recent paper, Liu et al. (73), in which

three different proteins were all found to function redundantly in both uncoating and DNA synthesis. How exactly this system facilitates either of these two processes, however, is not yet understood.

2.3.2 Origin of Replication

Until recently, no origin(s) of replication had been found. It had been theorized that POXVs perhaps do not need specific origin sequences, unlike nuclear DNA viruses (7). What was known, however, is that in cells infected with Shope fibroma virus (SFV) or VACV, any circular DNA molecule got replicated. Additionally, origin-independent plasmid replication takes place in viral factories and requires every protein necessary for genome replication, confirming the specificity of this activity. In experiments studying linear DNA molecules containing VACV hairpin ends, the presence of the terminal 200 bp resulted in enhanced replication. This terminal region contains the concatemer resolution sequence, but the mechanism of this enhancement is currently unknown (40). A recent paper, by Senkevich et al. (74), shed some light on this process by finding that origins possessing a prominent initiation point mapped to a sequence within only one of the isomeric hairpin loops at the end of a mature genome. Interestingly, this same sequence was also located within all the concatemeric junctions of replication intermediates. Given that these two sequences are identical, the group was unable to discriminate between the two. What they did find, however, was that their analysis was consistent with VACV DNA synthesis starting at a replication fork at an origin found within the terminal hairpin of a mature monomeric genome following infection, during the initial round of infection, and then subsequently near the center of the concatemeric junctions of the

replication intermediates that form during VACV DNA replication. What may represent the main start site of VACV DNA replication corresponds to a spot 35 nt from the apex of the hairpin. Additionally, several potential minor start sites were also detected in both directions. The finding of one prominent start site led to the conclusion that VACV DNA replication starts primarily on one end or is asymmetric. Also, its location within the hairpin loop is not surprising given that early experiments suggested the genome ends were the sites of initiation and that the loop is the most AT-rich region of the viral genome (74).

2.3.3 DNA Replication Model

Two possible models have been proposed for how POXVs replicate their genomes: one like the rolling hairpin strand displacement mechanism proposed as the method of replication for single-stranded Parvoviruses; or one involving RNA priming and semi-discontinuous DNA synthesis at replication forks, which is more conventional than the former model. The first model is supported by the fact that the POXV genome has a unique terminal structure and a high molecular weight, along with the presence of junction fragments and evidence suggesting that nicking and initiation occur near the end of the molecule. For this model to work, a hypothetical nick is placed at one or both ends of the genome, providing a free 3'-end for priming replication. The replicated DNA strand then folds back on itself and the replication complex copies the rest of the genome. Concatemer junctions form after the complex goes through the hairpin. Very large, branched concatemers arise when new rounds of replication are initiated before these structures are resolved. Once late-stage transcription begins, unit-length genomes are resolved for packaging into viral

progeny and the incompletely base-paired terminal hairpins are regenerated, with inverted and complementary sequences intact. This model has not been supported yet as neither the nicking site nor the essential nicking enzyme needed to begin DNA replication has been identified (40). The alternative discontinuous model was supported by early reports of VACV DNA covalently linked to RNA and the chasing of DNA into larger molecules, pointing to lagging strand synthesis as the mechanism (40, 74). The problem with this model is that, until recently, a virally-encoded RNA primase had not been discovered and the virally-encoded DNA ligase was not essential for viral replication. This model is now supported by the recent discovery of the long-sought viral RNA primase (75) and by the fact that the host DNA ligase can be substituted for the viral ligase when it is absent (76). More support for this model was gained after a group used deep sequencing to map the origins of replication. These findings were consistent with the early EM and biochemical studies looking at VACV DNA replication, along with the need for primase and ligase activities (74). The presence of branched DNA molecules in this model can be explained by what has been seen in phage T4, in which recombination-dependent DNA replication involves invasion of duplex DNA by the 3' OH of ssDNA. It is also entirely possible that POXVs use multiple mechanisms to successfully replicate their large genomes, as these mechanisms are not mutually exclusive (40), although this scenario seems far less likely now due to recent findings (74).

2.3.4 Enzymes Involved in DNA Precursor Metabolism

Some poxviruses encode the enzymes involved in DNA precursor metabolism, a trait which comes in handy when attempting to replicate in cells with

suboptimal precursor pools. These enzymes include a thymidine kinase (TK), thymidylate kinase, ribonucleotide reductase (7, 77), dUTPase, and an incomplete guanylate kinase (7), all of which are found in OPXVs. The incomplete guanylate kinase may indicate that an intact one exists in some other POXVs. Leporipoxviruses lack genes encoding the large subunit of ribonucleotide reductase, thymidylate kinase, and remnants of the guanylate kinase. MOCV lacks all these genes, which may contribute to its limited host range. FWPV encodes a protein related to human deoxycytidine kinase, which has yet to be found in any other Poxvirus. Entemopoxvirus *Melanoplus sanguinipes* does not encode any of these genes, but it does have a thymidylate synthetase homolog. A homologous TK gene has been found in other Entemopoxviruses however (40). The TKs encoded by POXVs are all 20-25 kDa in size and have 35-70% amino acid identity with their corresponding eukaryotic enzymes. Interestingly, they are not related in sequence to the pyrimidine kinase of herpesviruses. The VACV TK is, however, quite similar in structure to the human TK. The differences between the two relate to their associations with dTTP and are subtle enough that this enzyme could be a potential target for antivirals. The TK gene possesses an early promoter, while the enzyme exists as a tetramer, has adenosine triphosphate (ATP) and Mg2+ binding domains, and is susceptible to feedback inhibition by dTDP or dTTP. Even though TK is not essential for virus growth in cell culture, deletion mutants are severely attenuated in animal models (40).

The thymidylate kinase encoded by VACV is 23 kDa and can complement *Saccharomyces cerevisiae* mutants deficient in the homologous enzyme. This kinase catalyzes the next step in TMP metabolism, and like the TK gene, contains an early

promoter and is not required for replication in cell culture. The VACV thymidylate kinase has 42% sequence identity with its human homolog, but differs in substantial ways, such as having broader substrate specificity (40).

The ribonucleotide reductase is made shortly after VACV infects a new cell. This enzyme converts ribonucleoside diphosphates into deoxyribonucleoside diphosphates. It exists as a heterodimer, with the small subunit containing the catalytic domain and the large subunit being the regulatory subunit. Each subunit closely resembles its eukaryotic counterpart in terms of function and structure (70-80% identity). The enzyme's catalytic activity is inhibited by hydroxyurea preventing DNA replication. Interestingly, drug-resistant mutants generate direct tandem repeats of the gene coding for the catalytic (small) subunit. Mutating the regulatory (large) subunit prevents induced enzyme activity which does not affect replication in cell culture but mildly attenuates the virus in a mouse model. This virally-encoded heterodimer is only present in OPXVs, as many ChPVs only have the large subunit, which then must form a complex with the host small subunit in order to form an active enzyme (40).

The VACV dUTPase hydrolyzes dUTP into dUMP, which is an intermediate in the biosynthesis of TTP. Its activity may also minimize dUTP incorporation into DNA. Like the previous three enzymes, the dUTPase is made during the early phase of gene expression and is not required for the virus to replicate, in dividing cells at least. Its absence or mutation is felt more in quiescent cells, especially if the viral uracil DNA glycosylase is missing. Like the viral TK, the VACV dUTPase closely resembles the structure of the human homolog. Also like the viral TK, subtle

differences exist between the two, allowing for the possibility that this enzyme could be exploited as an antiviral target (40).

2.3.5 Viral Proteins Involved in DNA Replication

VACV encodes its own DNA polymerase (DNAP) (E9), complete with 3` exonuclease activity, a nucleic-acid-independent nucleoside triphosphatase (NTPase) (D5), a serine/threonine protein kinase (B1), a uracil DNA glycosylase (UDG) (D4), a DNA processivity factor (A20), and a functional ATP-dependent DNA ligase. The DNAP, serine/threonine protein kinase, UDG, NTPase, and DNA processivity factor are all essential for VACV DNA replication (7). The 117-kDa viral DNA polymerase (78) is homologous to other eukaryotic and viral DNAPs (40). Information regarding the active site has been gathered from codon substitutions at certain residues which confer resistance to inhibitors of DNA synthesis (40). This enzyme has been one of many viral proteins viewed as a potential antiviral target. One such compound, Cidofovir, even made it to the clinical stage of testing. This acyclic nucleoside analog, and its oral derivatives (78), get incorporated into the growing DNA strand, inhibiting the 5`-to-3` chain extension and 3`-to-5` exonuclease activities of the viral DNAP (40, 78).

The B1 kinase is a 35-kDa enzyme which is expressed early in infection, packaged in virions, and is required for VACV DNA replication. A host cytoplasmic protein known as barrier to autointegration factor, or BAF, is a potent inhibitor of POXV DNA replication. It has now been shown that B1-mediated phosphorylation of BAF blocks its DNA-binding activity. A ts VACV mutant defective in B1 kinase activity can be rescued by two human B1 homologs. B1 may have additional

substrates that are involved in immune defense activities (40) and translation of viral transcripts (5). Another known substrate for B1 is the early protein H5. Direct interactions between these two proteins were demonstrated via the yeast two-hybrid (Y2H) system. H5 seems to be a versatile protein as it reportedly has roles in DNA replication, transcription, mRNA processing, and morphogenesis (40).

D4, encodes a functional UDG, similar to the one found in SFV. These enzymes are involved in DNA repair, removing uracil residues that have been introduced into the DNA through either misincorporation of dUTP or cytosine deamination. VACV mutants with an enzymatically inactive UDG show attenuation in a mouse model, demonstrating that the repair function is beneficial for replication and virulence. This benefit is highlighted by the fact that the catalytic site is conserved in all POXV orthologs. Outside of POXVs, the POXV UDG only has about 20% sequence identity with non-POXV homologs, in addition to possessing different secondary and tertiary structures (40). D4 also plays a role as a processivity subunit of the viral DNAP, a function which does not require its enzymatic activity (79).

The D5R gene encodes a 90-kDa enzyme with both NTPase and primase activities. Mutations in the active site of either region are unable to complement a conditional lethal temperature-sensitive (ts) mutant. D5 is needed for viral replication, though it seems this is independent of its glycosylase activity (40). The requirement for D5 in DNA synthesis is emphasized by the fact that it is conserved across all POXVs, with more diverged orthologs found in members of all other families of the NCLDVs (75). D5 and its orthologs possess a motif found in the archaeo-eukaryotic

primase superfamily (40, 75). The discovery of a virally-encoded primase has important implications regarding the mechanism of genome replication and also provides an additional potential therapeutic target (75). One trait that sets D5 apart from the other viral DNA synthesis proteins is that it is currently the only one involved in uncoating of the viral core (73).

Another protein involved in DNA replication is the A20 protein, which has no non-POXV homologs. The role of A20 was first hinted at thanks to Y2H analysis, which showed an interaction between it and the viral DNA replication proteins D4, D5, and H5. Additional Y2H studies demonstrated that non-overlapping regions of A20 were responsible for these separate interactions, implying that these four proteins can interact simultaneously to form a multicomponent complex. These interactions were then confirmed via co-immunoprecipitation. Targeted mutagenesis showed that A20 was involved in VACV DNA replication, specifically regarding processivity when A20 is complexed with D4 and the DNAP. Another protein, which was found to be essential for VACV replication, is the 34-kDa phosphoprotein encoded by the I3L gene. This protein forms octameric complexes on ssDNA and is found in punctate cytoplasmic inclusions containing parental VACV DNA. All five of these proteins are expressed early in infection and were identified using complementation groups of ts/conditional lethal mutants that were impaired with regard to DNA synthesis (40), thus defining them as being essential for DNA replication (72, 75).

VACV encodes its own functional ATP-dependent DNA ligase (A50). It was initially found to be non-essential for replication in cell culture. It was then discovered that cellular DNA ligase I specifically was recruited to viral factories in

ligase deletion mutants. This means that some ligase activity is required for VACV replication. The viral ligase still has a significant role though, as its absence imparts sensitivity to DNA damaging agents and is important for virulence. Furthermore, encoding its own ligase allows VACV to enhance its replication in resting cells by boosting DNA synthesis. This is especially evident when ligase-deficient mutants are used to infect resting primary cells. Replication is greatly reduced and delayed, which correlates with low initial levels of cellular ligase I and subsequent viral induction and localization of the host ligase in viral factories (40, 76). Photolyases also play an important role in DNA replication by protecting DNA from ultraviolet (UV) radiation via excision of cyclobutane pyrimidine dimers. These are found in Avipoxviruses, Leporipoxviruses, and Entomopoxviruses and are likely important for POXVs which are transmitted through the environment versus from animal-to-animal (40).

2.3.6 Concatemer Resolution

During replication, POXV genomes form multimeric structures called concatemers, which must be resolved to generate unit-length genome molecules to be successfully packaged. The concatemer junctions consist of a precise duplex copy of the hairpin loop present at the ends of the mature DNA genome. The minimum sequence requirement needed for resolution to occur is two copies of T6-N7-9T/C-A3-T/A. This must be present in an inverted repeat orientation on either side of an extended double stranded copy of the hairpin loop (7). As mentioned above, this sequence also contains what is believed to be the main start site of VACV DNA replication. In the model proposed by Senkevich et al. (74), the junction would serve

as a bidirectional origin. Intriguingly, the minimum resolution sequence contains a functional late promoter, suggesting a link between concatemer resolution and transcription (40). Not surprisingly, the 35-nt start site located within this sequence is situated close to this well-characterized, strong late promoter. The transcript made by this promoter does not contain an extended ORF and its role, if any, in VACV replication is unclear. One theory is that the interaction of the vRNAP with this promoter helps to unwind the DNA duplex during initiation of replication (74). The sequence of the intervening region that is destined to form the hairpin loop, while not highly conserved, must be palindromic (59-61) and no more than 200 bp long (40, 62).

Concatemer resolution occurs quickly and independently of virion assembly, as junctions only accumulate if post-replicative gene expression is inhibited (40), specifically late gene expression (80). Junctions contain an inverted repetition, which, in supercoiled plasmids, makes a cruciform structure resembling a four-way Holliday junction (HJ). This led to the conclusion that the putative concatemer resolving enzyme would be an HJ resolvase. This was confirmed when the motifs and structural elements critical for the activity of the E. coli RuvC HJ resolvase were discovered in ORFs conserved in all POXV genomes. When this homolog, A22R in VACV, was expressed as a recombinant protein in E. coli, it was found to specifically resolve HJs, experimentally confirming what the bioinformatic analysis revealed. Similar to RuvC, A22 exists as a dimer in solution and when bound to HJ structures, but it possesses a lower degree of sequence specificity with regard to cleavage (81). An inducible null mutant was unable to process concatemers into unit-length

genomes with hairpin ends under non-permissive conditions, leading to the conclusion that the enzyme is required for resolution, though additional proteins may assist in determining site specificity (40, 82). This mutant displayed reduced late-stage DNA replication and inhibited concatemer junction cleavage. As such, most of the newly-replicated DNA remained in a branched or concatemeric form. It also had a block in virion morphogenesis at an early step in the process, yet protein synthesis was not affected (82). The requirement for late gene expression to occur in order for concatemer resolution to happen (80, 81) made sense when it was determined that the HJ resolvase is expressed late in infection. This enzyme also happens to be packaged into virus particles (82). A similar phenotype was observed in a paper published by Katsafanas et al. (77). What this paper found was that by depleting the pyrimidine nucleotide pool, they could reduce VACV replication by 3 logs. Specifically, even though viral DNA accumulation was only reduced by 60%, very little of it ended up being converted into genome-length molecules. They also noticed a drastic decrease in post-replicative gene expression (77), which would explain the DNA processing phenotype since the HJ resolvase is expressed after DNA replication takes place (82).

Type I DNA topoisomerases are another class of enzymes capable of resolving certain DNA structures. They form covalent links with DNA and relieve supercoils during replication, transcription, and recombination. POXVs encode their own topoisomerase. The POXV version of this enzyme has an unusual trait, however, in that it exhibits some sequence specificity. While its roles *in vivo* still need to be fully illuminated, it has been characterized *in vitro* to carry out a multitude of reactions including strand transfer, transesterification, recombination, and cleavage

and ligation of a variety of DNA structures, one of which is a HJ. Mutants in which this gene is knocked out are still able to undergo DNA synthesis and concatemer resolution. There is evidence, however, that this enzyme plays an important role in transcription. This makes sense upon realizing that the gene is regulated by a late promoter, which is more in line with the enzyme having a role in transcription or DNA processing rather than replication (40, 83).

2.3.7 Homologous Recombination

Recombination occurs quite often within POXV-infected cells. Such frequent occurrence, specifically between the terminal sequences of POXV DNA, may explain the variation in the number of tandem repeats found in addition to translocations and mirror image deletions. Recombination can have much more significant impacts on POXVs though. For instance, it has apparently taken place naturally between individual Capripoxviruses and between SFV, which produces benign fibromas in rabbits, and MYXV, the agent of Myxomatosis, to form malignant rabbit fibroma virus. Perhaps even more extraordinary is the fact that both field and vaccine strains of FWPV carry a near full-length and seemingly infectious integrated avian retrovirus genome. Even though the precise mechanism of recombination is not fully understood, what is known is that the process does not require post-replicative gene products and there seems to be a strong connection between recombination and replication. Evidence suggests that the DNA polymerase is directly involved in recombination and that the 3` exonuclease and DNA joining activities are both utilized (40, 84).

G5, an early protein conserved in all POXVs, belongs to the FEN1 family of exo/endonucleases. G5 deletion mutants are severely impaired and there is a 100-fold reduction in infectious virus yield. The virions produced seemingly contain the normal complement of proteins but have a spherical shape, instead of the characteristic brick shape, and contain little to no DNA (40, 84). Similarly-shaped virions have been observed with A32, an ATPase (85), and I6, a telomere-binding protein, mutants, both of which are required for DNA packaging. G5 itself may also be required, as it is normally present in viral cores (84). Additionally, even though the amount of viral DNA produced by these mutants is similar to that made by the wild-type virus, the mean size is about one-fourth the full-length genome (40, 84), suggesting a role in full-size genome formation (72). Most of the DNA made in G5 deletion-infected cells is not packaged, even though HJs were resolved, forming mature termini. Senkevich et al. postulated that there may be a length requirement for packaging DNA into virions or that mature termini are required on both ends of the genome (84). Experiments using transfected plasmids have shown that G5 is required for double-strand break repair via homologous recombination, which led to the assumption that it played a similar role in VACV genome replication (40, 84). This kind of DNA damage commonly occurs at replication forks, which further supports the semi-discontinuous model of POXV DNA replication (40).

Recombination can also occur between viral genomic DNA and transfected subgenomic DNA fragments, a fact which has been exploited to map and construct mutations as well as insert genes for expression. The viral genomes rapidly eliminate direct repeats via the formation of intra- and intermolecular recombination products.

Both single- and double-crossover products resulting from recombination between transfected plasmids and viral genomes, as well as both intra- and intermolecular plasmid or bacteriophage DNA recombinants have been detected in POXV-infected cells. All ChPVs except Avipoxviruses encode homologs of serine recombinases. Only the Crocodilepox virus (CRV) homolog, however, contains all the amino acids which make up the active site, and deletion of the VACV homolog, F16 (79), has no effect on replication in cell culture (40). In fact, the role of F16 in virus reproduction is unclear, although it was found to unexpectedly localize to the nucleoli of infected cells (79).

2.3.8 Viral DNA-Associated Proteins

In total, there are currently 10 viral proteins known to be involved in DNA replication, with another 4 involved in precursor metabolism, and 2 more involved in packaging DNA into newly formed viral particles. The viral proteins required for DNA replication, as mentioned above, are the DNAP (E9), a processivity factor (A20), a UDG (D4), a multifunctional scaffold protein (H5), and an NTPase (primase/helicase) (D5). The additional viral proteins involved in this process are the ssDNA binding protein (I3), the DNA ligase (A50), the serine-threonine protein kinase (B1), a FEN-1 homologous nuclease (G5), and an HJ resolvase (A22). These are all early proteins, made before DNA synthesis occurs, with the lone exception being the resolvase. It is not required for DNA synthesis, as it is needed to process concatemeric replicative intermediates into unit-length genomes with hairpin ends. While this list would lead one to believe that VACV does not require additional factors for this process, there is doubt as to whether other viral or host proteins may

be involved somehow, perhaps by enhancing replication or providing some missing functions. Studies have found evidence for roles of the cellular topoisomerases IIA and IIB (TOP2A and TOP2B), DNA ligase I (76) and DNA damage response proteins, the ssDNA binding protein RPA, and the DNA clamp protein proliferating cell nuclear antigen (PCNA). Additionally, there are host defense molecules which act in opposition to VACV infection, particularly DNA replication. These are BAF1, which acts as a negative regulator of VACV genome replication, and DNA-PK, which is a sensor of cytoplasmic DNA. While some of these host proteins have been demonstrated to localize in viral factories, direct association with the VACV genome *in vivo* has yet to be observed (72).

To this end, Senkevich et al. (72) undertook a study to address these gaps in knowledge regarding the VACV DNA replication machinery. What they found was the complete set of viral replication proteins associated with the nascent DNA. This included the seven proteins necessary for DNA synthesis (E9, A20, D4, H5, D5, I3 and A50), plus G5, I1, and B1. Some of these interactions may have been direct, while others may have been mediated by another protein. For instance, D4 can bind DNA and forms a holoenzyme complex with A20, which has not demonstrated DNA-binding activity, and E9. H5 is also able to bind DNA and interacts with A20 as well, and B1. The three most abundant viral proteins detected were H5, the I1 telomere-binding protein, which is required for morphogenesis instead of DNA replication, and I3, which is consistent with their abilities to directly bind DNA, as well as their abundant transcript levels at 4h.

Host proteins found associated with the nascent DNA include TOP2A and TOP2B, PCNA, several additional host DNA replication/repair proteins like HMGA1, TOP1, PARP1, and RUVBL2, and DDX3X. The topoisomerases were only recovered from cells infected with wild-type virus, and not from cells infected with the ligase-deficient virus, supporting the findings of another group. Despite this though, the VACV DNA ligase mutant was still capable of replicating, calling into question what role(s) the type II topoisomerases play in VACV DNA replication. PCNA increases the processivity of the DNA polymerase in eukaryotic DNA replication. VACV, however, already encodes two proteins which serve a similar function, D4 and A20. Other studies have provided evidence for a role of PCNA, though. One report found that two human PCNA-associated proteins were inhibitors of a host range mutant (86) while another reported that T2AA, a PCNA-specific inhibitor, and siRNAs specific for PCNA both inhibited VACV DNA synthesis (72). In light of these reports, Senkevich et al. (72) found that T2AA did indeed inhibit VACV DNA replication in a concentration-dependent manner. Such results will likely lead to more studies regarding the role of PCNA in VACV DNA replication. DDX3X is a DEAD box helicase, which has been reported to be a target of the VACV K7 protein and an activator of IRF3. The specificity of its association with the nascent DNA is questionable though. There is no evidence yet to support the remaining host proteins detected having essential roles in VACV DNA replication.

Surprisingly, the cellular DNA ligase I was not detected in this study with the viral ligase-deficient mutant. This host protein in essential for VACV DNA replication when the viral ligase is absent and has been observed localizing in the

viral factories. Studies that have implicated DNA damage response proteins (ATR, CHK1, Rhino, INTS7, and TOPBP1) and the DNA replication protein RPA were not supported by this study, as none of these proteins were detected. It had been proposed by another group that RPA served as the replicative ssDNA binding protein, not I3 (72). This proposal was not supported by the findings of Senkevich et al. (72), as I3 was one of the most abundant proteins detected and none of the RPA subunits were detected (72).

2.4 Virion Morphogenesis, Maturation, and Egress

Virion assembly begins in structures known as the viral factories, which are circumscribed, granular, electron-dense areas of the cytoplasm. The first distinct morphological structures that can be visualized are crescents and immature virions (IVs). The latter structures contain a membrane with spicules on the convex surface and granular material in the concavity. Deep-etch EM revealed the spicule layer to be a honeycomb lattice viewed on edge. The mechanism of IV formation and the source of its membranes remained a conundrum for many years, with many hypotheses being put forth. Some evidence suggested that the viral membrane formed de novo, while another proposed that it was derived from a cellular organelle. There was evidence suggesting that the intermediate compartment between endoplasmic reticulum and Golgi apparatus (ERGIC) was not the source of the IV membrane. Additionally, no ER or ERGIC protein had been found to co-localize with the IV membrane and no viral proteins identified as being incorporated into purified MVs possessed any signature of ER translocation, such as signal peptide cleavage or glycosylation (7). In spite of this, the lipid profile of purified MVs is interpreted as

being consistent with the MV membrane originating from the intermediate compartment or cis-Golgi network (47). It had been found that a heterologous signal peptide fused to the N-terminal region of a VACV membrane protein was cleaved. The truncated protein then localized in IVs and MVs, providing evidence for at least a functional pathway between the ER and viral membranes (40). In the intervening years, however, it has since been determined that the ER is indeed the source of the viral membranes, through the use of deletion viruses missing proteins critical for viral membrane assembly (6, 53-55, 87, 88).

2.4.1 Crescent and IV Formation

Viral proteins required for IV formation, F10 kinase, H5, G5, A11 (7), H7, L2 (40), and I2 (66) have been identified through the use of conditional lethal VACV mutants (7) and deletion mutants in conjunction with a complementing cell line (53, 66). Interestingly enough, none of these proteins make up the IV membrane (7). Both H5 and G5 have other roles in viral replication, so they may have indirect roles in crescent formation (40). In addition to being required for IV formation, F10, A11, H7, and L2, along with A17, A14, and A6, are all essential for crescent formation. Without even one of these proteins, dense masses of viroplasm along with vesicles and tubules often accumulate in the cytoplasm instead of the characteristic crescents and IVs (6, 54, 87, 88).

A6 is a 43-kDa protein that is conserved in all ChPVs. It has no predicted TM domain, is expressed post-replicatively, is packaged to a minor extent in MV cores, and has been shown to associate with the A11 protein. H7 also has no predicted TM domain and is conserved in all ChPVs. It is a 40-kDa protein, is not detected in

highly-purified MVs, and is not strongly retained in viral factories following synthesis (87).

L2, unlike the other proteins needed for IV formation, is made during the early phase of infection, contains a TM domain, associates predominantly with the ER throughout the cytoplasm (54), and can be found near the growing edge of crescents (40, 52). All these characteristics set it apart from other proteins involved in this process (54). Regarding its topology, the N-terminal end faces the cytoplasmic side of the ER while the C-terminal end faces the luminal side (89). L2 is conserved across all ChPVs (90), hinting at how essential it is for VACV and consistent with its important role in virion morphogenesis. This made characterizing the protein difficult, requiring either the construction of an inducible mutant (90) or a deletion mutant, which could only be isolated with the help of a complementing cell line (53). When L2 expression is blocked, proteolytic cleavage of the major core proteins and A17, an essential component of the IV membrane, fails to occur. Such a phenotype suggests that the block in virion morphogenesis happens early in the process. Repression of both A11 and A17 yield similar phenotypes (52). Additionally, repression of any of these three proteins drastically decreases the stability, and thus the amount, of a certain subset of viral proteins (52), 25 to be exact (53), including members of the EFC complex (52, 53). As such, it appears the role of L2 is two-fold. It is required for the formation or elongation of crescent membranes (90), which it does by recruiting ER and modulating its transformation into viral membranes in the vicinity of the viroplasm. This recruitment in turn prevents the degradation of this

subset of viral proteins which are entirely dependent on viral membranes for stability (52, 53, 87).

A11 is made during the post-replicative phase of viral gene expression, like other virion morphogenesis proteins, localizes in cytoplasmic viral factories, self-associates to form dimers or higher-order structures, is phosphorylated independent of the F10 kinase, and is absent or not found in significant amounts with purified MVs. This last trait it shares with both H7 and L2. In fact, A11 shares many features with L2. Both are present at the edges of crescent viral membranes, over the course of a normal infection, and are found on or near aberrant membranes in mutant virus-infected cells. Both proteins are tail-anchored and associate, post-translationally, with microsomal membranes, which was demonstrated by *in vitro* transcription and translation experiments. Despite these similarities, A11 possesses some features which distinguish it from L2. For instance, A11 co-localizes with the ER, but only in viral factories. This contrasts with L2, which also co-localizes with the ER, but throughout the cytoplasm. In uninfected cells, L2 still co-localizes with the ER and was found to associate with membranes. On the other hand, A11 does not, even when L2, A6, or H7 are co-expressed. When A11 expression is knocked down, large, dense bodies formed, like what is seen when working with H7 and A6 mutants. IV-like (IV-L) structures are also observed trapped in the ER when A11 expression is blocked. This phenotype had only been previously seen with an L2-deletion mutant, highlighting yet another common feature shared by these two proteins (54, 87).

A previously uncharacterized protein was also found to be vital for virion morphogenesis. Referred to as A30.5, this ORF was revealed to not be annotated in

the genome sequences of many POXVs, even though ORFs corresponding to the same location have been found in representatives of all ChPVs. The lack of annotation may be due to its small size, at just 42 aa in length (55). Its well-conserved nature, however, would imply that it is important in the virus lifecycle. A30.5 was found to interact with L2, associate with the ER, participate in biogenesis of the ER, and be required for replication (55, 87). Despite its association with L2, its topology differs from that of its interaction partner. It was discovered that both the N- and C- termini face the cytoplasm (89). The study, by Maruri-Avidal et al. (55), characterizing this protein also found evidence of direct formation of the crescent membranes from the ER, in addition to finding that D13 localized to the luminal side of the ER membrane. This led to the conclusion that the outer surface of the MV is in fact derived from the luminal side of the ER. The report that crescent membranes were directly connected had been confirmed by transmission electron microscopy (TEM), but not by electron tomography (ET), until recently, when the continuity between these two structures was unambiguously demonstrated (6).

2.4.2 VMAPs

Using null mutants, multiple proteins were found to be involved in a similar or the same step of virion morphogenesis as L2. This group of proteins has been dubbed VMAPs, or viral membrane assembly proteins. The proteins designated as such are L2, A30.5, A11, H7, and A6 (6, 55, 66, 87, 88). These five proteins are scarcely found or completely absent from MVs, despite their importance for IV formation. If expression of any of the VMAPs is blocked or drastically inhibited, electron-dense inclusions form in place of normal-looking IVs. That said, there are differences seen

in the phenotypes of VMAP-knockout mutants. The structures seen with L2 and A30.5 deletion viruses are virtually identical, which was not surprising given that the two proteins form a heterodimer. Surprisingly, the A6 deletion mutant also formed identical structures, perhaps hinting at functional and physical associations with L2 and A30.5. While indistinguishable structures were also seen with the H7 deletion virus, there were differences present. In these infected cells, there was more extensive ER wrapping plus additional particles which looked like hooks and double-wrapped IVs. Such structures were also observed in cells infected with an A11 deletion virus. ET revealed, however, that these also represented incomplete spherical or crescent structures which had connections to the ER. As such, all five VMAPs share a common feature: deletion of any results in continuity between the viral and ER membranes. It is currently unknown, though, whether higher-order structures with all five VMAPs exist. This study, by Weisberg et al. (6), also demonstrated that A17 and D13, which are known interaction partners, both have membrane-remodeling properties. Taken together, these data were used to propose a model for viral membrane formation from the ER (**Fig. 2.2**).

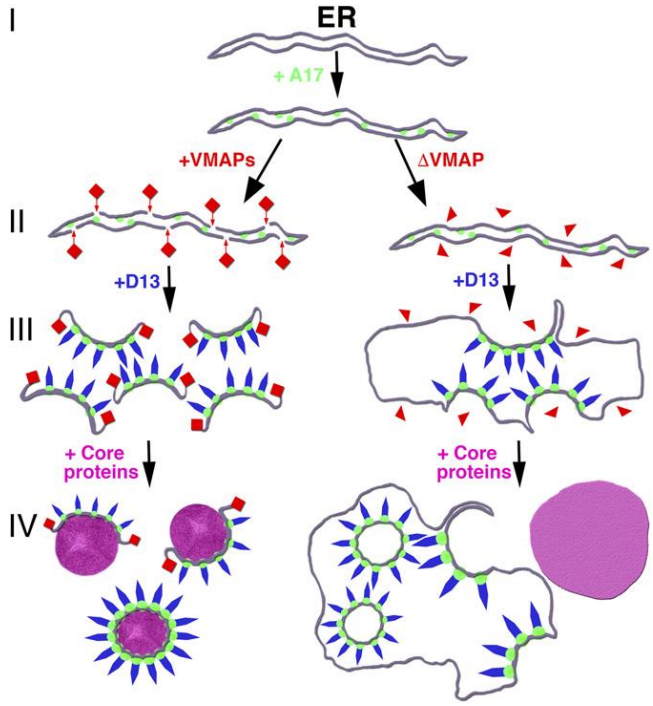

Figure 2.2. Model of viral membrane formation from ER. The first step (*I*) consists of modification of the ER by insertion of the A17 transmembrane protein, represented by green spheres. In the presence of all five VMAPs, represented by red rectangles, membrane scission occurs, and the edges of the sheets are stabilized, as shown on the *Left* (*II*). D13, represented by blue shapes, associates with the N terminus of A17 to increase curvature-forming crescent structures (*III*), which extend around core proteins (purple) to form the spherical IVs (*IV*). VMAPs missing one component (ΔVMAP), represented by red triangles, are unable to induce membrane scissions, the crescents remain attached to the ER, and empty IV-L particles bud into the lumen, as shown on the *Right*. The core proteins form dense inclusion aggregates (purple) outside of the ER. Figure 6 from Weisberg, A.S., et al. Enigmatic origin of the poxvirus membrane from the endoplasmic reticulum shown by 3D imaging of vaccinia virus assembly mutants. Proc Natl Acad Sci USA, 2017. 114(51): p E11001-E11009. (6). Reprinted with permission.

In this model, the ER is first modified via insertion of A17. The ER membrane is then broken up and its loose ends stabilized by the VMAPs in conjunction with other viral proteins or perhaps even cellular proteins. D13 then binds to A17, providing additional curvature of the viral structures. These crescents then extend around core proteins to form spherical IVs. In the absence of even one VMAP, ER membrane scission does not take place. This results in the crescents remaining attached to the ER and the budding of empty IV-L particles into the lumen. The core proteins then form dense inclusion aggregates outside of the ER (6, 87, 88).

As stated earlier, I2 is a small protein, at only 72 aa long and with a mass of 8.4 kDa. It possesses a C-terminal TM domain, is made after VACV DNA replication, is associated with purified MVs, and is conserved in all ChPVs. It was found that virions lacking I2 are unable to enter cells, thus greatly reducing infectivity (66). The Hyun et al. (66) study set out to investigate the initial report from another group regarding this defect in viral entry by determining which step of replication or entry was affected by the absence of I2. What they found was that, while an I2 deletion mutant was unable to enter cells, the primary block was, unexpectedly, in virion morphogenesis, specifically after the formation of IVs. This resulted in an accumulation of dense, spherical particles with poorly formed cores instead of the characteristic brick-shaped MVs with well-defined core structures. The IVs that formed appeared normal and contained normal amounts of most core and membrane proteins. However, during maturation, the incomplete MVs that formed had a drastic reduction in the amounts of the EFC proteins, had more D13 remaining, and greater amounts of unprocessed A3 and A17 proteins. This block in disassembling the D13

scaffold was a phenotype that had previously been observed when I7 expression was inhibited. I7 is the protease responsible for cleaving A17, in addition to several core proteins. Another intriguing find was that there were reduced amounts of I7 found in cytoplasmic extracts of non-complementing cells infected with the I2 deletion virus, perhaps hinting at a relationship between I2 and I7. The failure to remove the D13 scaffold prevents the insertion of the EFC proteins into viral membranes and their hydrophobicity causes them to be unstable without the membranes, thus leading to their enhanced degradation. They concluded that I2 was required for virion morphogenesis, release of the D13 scaffold, and the association the EFC proteins with viral membranes. This overall phenotype, involving a block in morphogenesis, inability to form and accumulation of membranes free of D13, and a drastic reduction of EFC proteins, is also seen in VMAP mutants as it turns out. The difference though, is that the VMAP mutants encounter their blocks earlier in morphogenesis, prior to the formation of IVs, whereas I2 mutants actually progress to the IV stage of virion maturation (66).

2.4.3 Core Protein Association with IVs

Work with conditional lethal VACV mutants also revealed that seven core proteins are each required for the association of crescent membranes with the granular viroplasm. These core proteins are: A15, A30, D2, D3, G7, J1, and the F10 kinase. These seven proteins form a complex, presumably linked to one or more membrane proteins which have not yet been identified (7). Similar phenotypes are observed when expression of the core proteins A10 (7), E6, and F17 (also called F18) are repressed, with F17 being found to associate with the complex (40). F10 is important

because it phosphorylates A14 and A17, which are IV membrane proteins, and its kinase activity is critical for its function in morphogenesis (40). A14 and A17, the latter of which is a cellular reticulon-like protein (91), are vital for morphogenesis because when the synthesis of either protein is repressed, crescent membrane formation is reduced or abrogated and small vesicles accumulate. The fact that similar phenotypes are obtained with either one or the other supports evidence of interactions between these two proteins (40, 87), an interaction which has since been confirmed, although it is still not known whether it is a direct interaction (87).

2.4.4 D13 Scaffold and Rifampicin

The antibiotic rifampicin was discovered to negatively impact assembly, as its presence during infection leads to the accumulation of irregular viral membranes lacking the honeycomb lattice. Within minutes of the drug's removal, the single membrane bilayer becomes coated by the protein responsible for forming the honeycomb lattice, resulting in crescents being formed. This process occurs even in the presence of RNA or protein synthesis inhibitors, which led to the conclusion that rifampicin directly interferes with assembly. The protein responsible for forming the lattice and rifampicin resistance was found to be D13, which, when repressed, resulted in a similar morphological phenotype as when assembly is attempted in the presence of rifampicin (7). D13 forms trimers which come together to form the honeycomb lattice scaffold, leading to the spherical shape of VACV IVs. D13 interacts with the A17 N-terminal region to form the scaffold (7, 87). Duplication or overexpression of A17 confers rifampicin resistance to VACV (87). The existence of a domain shared between D13 and the capsid proteins of certain other lipid-

containing viruses suggests that the developmental stages of POXVs reflect evolution from an ancestor with an icosahedral capsid (7) into what is the present MV form. POXV retention of a capsid-like structure around the IV supports this by demonstrating that, like embryogenesis, morphogenesis recapitulates evolution. The predicted similarity of D13 with the capsids of other viruses and phage has been supported by structural studies (40, 87).

2.4.5 Genome Packaging

The next step in morphogenesis involves a dense mass of nucleoprotein entering the immature envelope, before being completely sealed. When assembly is interrupted prior to the formation of IVs, viral DNA concatemers are still processed normally and mature DNA accumulates in large crystalloid structures. It was also determined that, while concatemer resolution can take place in the absence of viral morphogenesis, DNA processing is required for morphogenesis to occur (7, 82). As mentioned earlier, when expression of the viral HJ resolvase is blocked, virion morphogenesis is blocked at an immature stage during this process (82). One protein found to play a crucial role in DNA packaging is A32, whose repression leads to a block in VACV genome packaging (7, 85). This activity was predicted based on its sequence similarity to the gene I of filamentous single-strand DNA bacteriophages and to the Iva2 gene of adenovirus, which are both ATPases involved in DNA packaging (7). A32 mutants do not go on to form mature particles with defined core structures, even though normal-looking crescents and IV particles can be seen beforehand. Very few of the particles have nucleoids and the result is an electron-dense, spherical particle that lacks DNA and has low infectivity and transcriptional

activity. Infection with these mutant viruses also led to the formation of large cytoplasmic DNA crystalloids, similar to what is seen when VACV infection is carried out in the presence of rifampicin (85).

Two other proteins which play important roles in genome packaging are I6 and A13, which are telomere-binding and viral membrane proteins, respectively. DNA packaging fails to occur under non-permissive conditions in cells infected with an I6 mutant, while conditional lethal A13 mutants display a defect in genome packaging and lead to the accumulation of DNA crystalloids associated with membrane. Mutation of I1, another telomere-binding protein, unlike results seen with I6 mutants, causes a defect at a later step in morphogenesis. It is not currently known how A32, I6, and A13 enable entry (7).

2.4.6 Packaging of the Transcriptional Complex

Components needed for early transcription must also be packaged into virions. These include RNA-associated protein of 94 kD (RAP94), VACV early transcription factor (VETF), the vRNAP, poly(A) polymerase, capping enzyme, topoisomerase, nucleoside triphosphate phosphohydrolase (NPH) I, and NPH II (7). A recent paper working with a panel of RAP94 mutants found that progeny virions produced by these viruses lacked nucleocapsids and had aberrant morphologies, which could be described as empty or collapsed in appearance (92). The mutant progeny contained wild-type levels of DNA but were selectively deficient in several transcriptional enzymes and thus had a reduced capacity to catalyze *in vitro* early transcription and reduced infectivity. These findings support the idea that the nucleocapsid is made up of the viral genome complexed with viral transcription enzymes and structural

proteins. Going one step further, the authors present the idea that the vRNAP and the associated transcriptional enzymes are in fact a structural component of the nucleocapsid, meaning that when these proteins fail to be encapsidated, the structural integrity of the nucleocapsid becomes compromised (92). Repressing the synthesis of VETF leads to accumulation of IVs (7). A32 mutants contain representative amounts of the vRNAP, VETF and several other enzymes, in addition to membrane proteins and cleaved and uncleaved core proteins. These findings suggested that the incorporation of these components is not tightly coupled with DNA packaging (85).

2.4.7 MV Formation

Once these previous stages have occurred, the spherical IV loses its lattice scaffold, D13, resulting in its conversion to a barrel-shaped particle, the MV. Disassembly of D13 is correlated with processing of A17 by the I7 protease. When I7 expression is blocked, D13 is retained on aberrant virus particles (40, 66). This transition process also involves the association of additional membrane proteins, and the processing and reorganization of core components. The transition from IVs to MVs is impaired when expression of A9, L1, H3 (7), or A6 (40) are knocked down. Defective-looking MVs are produced when some core proteins are mutated or their expression is blocked altogether (7). A recent study by Sivan et al. (93) found that knockdown of nuclear pore genes greatly inhibited VACV spread. There was no effect on viral entry, with only modest effects on early and late viral gene expression and viral DNA replication. The screen found that production of infectious virus was severely reduced, specifically beyond the IV stage of the virion morphogenesis process. This phenotype has also been observed in VACV infection of enucleated

cells. The likely reason for the defect found in this study is perturbations of nuclear transport. Whether or not disruption of this nuclear transport results in the inability of necessary cellular molecules to exit the nucleus or viral defense proteins to enter the nucleus is currently unknown, although the latter scenario seems more plausible.

2.4.8 Intramolecular Disulfide Bonds

Some of the MV membrane proteins contain intramolecular disulfide bonds in their cytoplasmic domains. Normally, disulfide bonds form in the ER. This oddity of the MV membrane proteins raised the possibility that POXVs encode novel oxidoreductases. This bore out upon confirmation that a unique cytoplasmic disulfide bond pathway was formed by three virally-encoded proteins: E10, A2.5, and G4. Repressing any of these three proteins blocks virion maturation (7). The structure of G4 has since been solved. It crystallized as a dimer, burying the Cys-X-X-Cys active site which may protect the reactive disulfides from being reduced in the cytoplasm (40). So far, nine viral membrane proteins have been identified containing intramolecular disulfides formed by this cytoplasmic redox system: L1, F9, and seven of the eight entry-fusion proteins. As of yet, no evidence suggests that this system forms the disulfide bonds of other membrane proteins or core proteins (40). This redox system is the only virally-encoded one discovered so far, in addition to being the only one known to operate on the cytoplasmic side of a membrane in any system. An ortholog of each protein involved in this redox system is found in every known POXV, emphasizing its ancient origin and requirement for POXV replication (7). The fact that the entry proteins have also been found in all known POXVs suggest their co-development (40).

2.4.9 Proteolytic Processing

Several VACV proteins undergo proteolytic processing during morphogenesis. These include A17, a membrane protein, and A3, A10, A12, L4, and G7, which are all core proteins. A17 processing is dependent upon the formation of the membrane component of the IV, while core protein processing will fail to occur if a later step in morphogenesis is blocked (40). Cleavage occurs at a consensus sequence in each case: AG_X. The protease mostly likely responsible for these cleavage events is I7, based on the observed phenotypes seen when studying I7 mutants and mutagenesis of the conserved cysteine protease active site. Another candidate is G1, which contains a motif found in some metalloproteases, HXXEH. Indeed, conditional lethal G1 mutants are blocked in morphogenesis, but the processing of the known membrane and core proteins is unaffected. Additionally, mutating the putative active site renders G1 incapable of rescuing a null mutant. Interestingly, G1 appears to undergo proteolytic processing, leading to the theory that it mediates self-cleavage (7).

2.4.10 MV Occlusion

The MVs of some ChPVs, such as CPXV, ECTV, Raccoonpox virus, FWPV (7), and Volepox virus (94), become occluded in a dense protein matrix within the cytoplasm, referred to as A-type cytoplasmic inclusion bodies or ATIs. They are distinct from the sites of viral replication and assembly, which are sometimes referred to as B-type inclusions (7). Some OPXVs such as VARV, VACV, HSPV and MPXV do not form ATIs but still retain a disrupted form of the gene, which may have evolutionary implications (94). The thought is that these ATIs are released into the

environment following degeneration of infected cells (7), which protects the enclosed MVs from environmental stresses (40), such as UV radiation, thus prolonging infectivity. For those POXVs that do not form these structures during infection, it seems they rely instead on the inherent stability of the MVs in the environment or direct animal-to-animal spread (94). The CPXV ATI protein is 160-kDa and may account for up to 4% of the total protein at late points of infection. Some VACV strains do encode a truncated version of a homologous protein, even though VACV does not form ATIs. The inclusion proteins of both CPXV and VACV are myristylated. Some CPXV mutants can form inclusions without virions, indicating a role for a protein specific to the MV known as 4C. The VACV homolog is A26. This protein acts as a bridge between the ATI protein and A27, which is tethered to MVs via A17. In the VACV strains that do encode the ATI homolog, it is also associated with A26. EVs do not contain either A26 or A25 (40). Entomopoxvirus virions also undergo occlusion where, following ingestion by a larval host, infectious particles may be released in the alkaline pH of the gut. Several Entomopoxviruses have these occlusion proteins, called spheroidin or spherulin. These homologs are all cysteine-rich, about 100 kDa in size, and lack homology to fusolin, the abundant 50-kDa spindle-body protein of Entomopoxviruses, the ATIs of ChPVs, and the polyhedrin protein of baculoviruses (7). Unlike MVs, neither the crescent nor IV membrane has been extensively purified and characterized, so their respective lipid compositions are currently unknown but they are assumed to be similar to that of the MV (47).

2.4.11 EV Formation

The other infectious form of the POXV virion, the EV, consists of an MV wrapped in an additional membrane. This extra membrane can be generated in one of two ways: an intracellular MV wrapping process, followed by exocytosis, or direct budding of the MV from the plasma membrane. FPXV on the other hand, exits from the cell primarily by budding (40). The former mechanism has been studied more extensively.

2.4.12 Wrapping of EVs

Some MVs are wrapped by two additional membranes, forming wrapped virions or WVs, which are also referred to as intracellular enveloped virions (IEVs) (47, 48), both of which are derived from virus-modified trans-Golgi or endosomal cisternae (7). With one exception, the phospholipid components of the WV are 2-3 times higher than in the MV (47). These extra membranes contain at least 8 viral proteins, A33, A34, A36, A56 hemagglutinin, B5, F12, F13 (7), and E2 (40), all of which contain a TM domain, with the exceptions of F12, F13, and E2. Of these eight proteins, A36 and F12 are exclusive to the outermost membrane, which is lost during exocytosis, and thus are not retained on the EV. The palmitoylation of cysteines 185 and 186 are essential for F13's association with the wrapping membrane (7), while F12 reportedly associates with E2 and A36 (40, 79). Deletion of any of the above genes results in a small plaque phenotype. The only exception is an A56 deletion mutant, which yields a syncytial phenotype (7) that may be due to its inability to bind to the EFC (40). The composition of the EV membrane is determined by complex interactions between A33, A34, and B5. A33 contains unique C-type lectin domains

that could be responsible for interactions with other viral or cellular proteins (40). Studies done with VACV mutants have shown that several proteins are required for efficient wrapping. The most severe effects are seen when the A27 MV protein, the F13 protein, and the B5 glycoprotein are either repressed or deleted. F13 is needed for the Golgi membrane localization of B5, in addition to having a putative phospholipase motif that is required for its role in wrapping and induction of post-Golgi vesicles (7). Another recent study by Sivan et al. (95) provided more insight into the wrapping process and viral egress, determining that F13 was transported to the sites of wrapping via retrograde transport. This is in contrast with B5, which is transported via the secretory, or anterograde, pathway. This novel use of retrograde transport, and the use of two different pathways, while seemingly inefficient, ensures that virions are not wrapped prematurely.

2.4.13 Intracellular Movement and Exocytosis

Within cells, WVs rely on microtubules for long-range movement, and not on actin polymerization as had been originally thought. The process of moving MVs from the sites of virion assembly to the sites of wrapping is dependent on microtubules, but the viral attachment protein is not currently known. There is thought that A27 is required for movement and wrapping, but this debate has not been conclusively settled. It has been demonstrated that the A36 protein interacts with the cargo-binding domain of the light chain of kinesin, a microtubule motor protein (7), which F12 reportedly had structural similarity to (40). This shared, functionally-relevant similarity between F12 and the tetratricropeptide repeats (TPR) of kinesin light chains (KLC) has since been disputed, however, by Yutin et al. (79). This study

revealed that the ChPV F12 protein is a derived family B DNAP in which the activities of both the polymerase and exonuclease domains have been abrogated due to mutational replacement of catalytic amino acid residues. Their analysis found no similarity to TPRs. Additionally, the presence of all-alpha TPRs is inconsistent with both the predicted structure of F12 and the alpha-beta nature of the DNAP domains. This paper also states that no quantitative evidence has been presented in support of this TPR similarity. Furthermore, TPRs have also been reported in both E2 and A36, both of which reportedly interact with F12. Yutin et al. (79) found no presence of this region in either ChPV protein. All three reportedly contribute to IEV maturation and intracellular motility (79). F12 also has a motor-binding motif vital for virion export (40, 79), via the recruitment of kinesin-1, which allows for movement of the IEVs along microtubules and that deletion of the purported TPRs in F12 abrogated kinesin binding (79). Yutin et al. (79) found that while their results do not necessarily conflict with those experimental observations, they do suggest that the interaction between kinesin and F12 is mediated by the derived DNAP domains.

When WVs reach the periphery of the cell, the outermost membrane fuses with the plasma membrane, liberating the EV, which now only has one more membrane than an MV (7). The migration of WVs through the dense cortical actin is mediated via the F11 protein prevention of RhoA signaling (40). The majority of EVs adhere to the cell's surface, while only a fraction of virions is found in the medium. The ratio of cell-associated to free virions is dependent on which VACV strain is being used (7, 48). This difference, at least in one case, appears to largely be the result of a single amino acid difference in the putative lectin-binding domain of

the A34 membrane protein (7, 48). Mutations in A33R and B5R have also been found to enhance release. The host Abl-family tyrosine kinases and phosphoinositide 5-phosphatase SHIP2 have also been implicated (40). EVs are important for virus dissemination, with cell-associated EVs being responsible for efficient cell-to-cell spread and plaque formation, in the context of a cell monolayer, and EVs in the medium providing long-range dissemination (7). While MVs are quite stable and are believed to be responsible for transmission between hosts, EVs, with their fragile outer membrane, are well-adapted to exiting an intact cell and spreading within a host (48).

2.4.14 Actin Tail Formation

Adherent EVs rely on their location at the tip of motile, actin-containing microvilli in order to efficiently spread from cell to cell (7, 48). When A33, A34, or A36 are not expressed, these actin-containing microvilli are not formed, yielding viruses that form small plaques. The nucleation of the actin tails is dependent on tyrosine phosphorylation of the A36 protein of OPXVs and the functional homologs found in other POXVs by Src- or Abl-family kinases (40). The phosphorylated form of A36 then interacts with the adaptor protein Nck, which leads to the recruitment of the Ena/VASP family member N-WASP to the site of actin assembly (7). This process of viral egress is one potential target for antiviral development, with one compound, ST-246, making it to clinical testing (40). This novel small molecule is an OPXV egress inhibitor, developed by SIGA Technologies (96), and blocks the wrapping of MVs and thus the formation of EVs by targeting the F13 protein (40).

2.5 Poxvirus Gene Expression and Transcriptional Regulation

POXV genes are expressed in a manner known as a cascade mechanism, i.e., the products of the preceding stage regulate the next stage. Specifically, POXVs have three stages of gene expression: early, intermediate, and late. Nearly all the enzymes needed for carrying out this cascade are encoded by the virus, including a vRNAP and stage-specific TFs. The core of the virion contains a complete early transcription system which allows for synthesis of early mRNA shortly after infection. This also provides an explanation for why purified POXV DNA is not infectious. Early mRNA encodes for enzymes and factors needed for synthesizing viral DNA and transcribing intermediate genes. These intermediate transcripts are translated into enzymes and factors needed for late gene expression. Late genes include structural proteins, used for making progeny virions, and the early transcription factors, which are packaged into these progeny virions along with the vRNAP. Thus far, all POXV genes consist of a continuous ORF and there has been no evidence to suggest that RNA splicing occurs (7).

VACV has 118 early genes (3, 97, 98) (**Fig. 2.3**), although some included in this number have both early and intermediate or early and late promoters, and so may be classified as early genes for the sake of simplicity (97, 98). A study also found that, at 4h in VACV-infected HeLa cells, 25-55% of the mRNAs were viral and by 7h 80-90% of the RNA hybridized to VACV DNA. The overall amount of polyadenylated RNA stays fairly constant, meaning that the change in proportion is due to the simultaneous degradation of host RNA and robust synthesis of viral RNA. This rapid decrease was seen in thousands of cellular mRNAs, indicating a global

effect. There is a total of 93 post-replicative genes (3, 97, 98). Similar to the early genes, some post-replicative genes have been categorized as intermediate even though they may also be expressed during the late phase of infection (98).

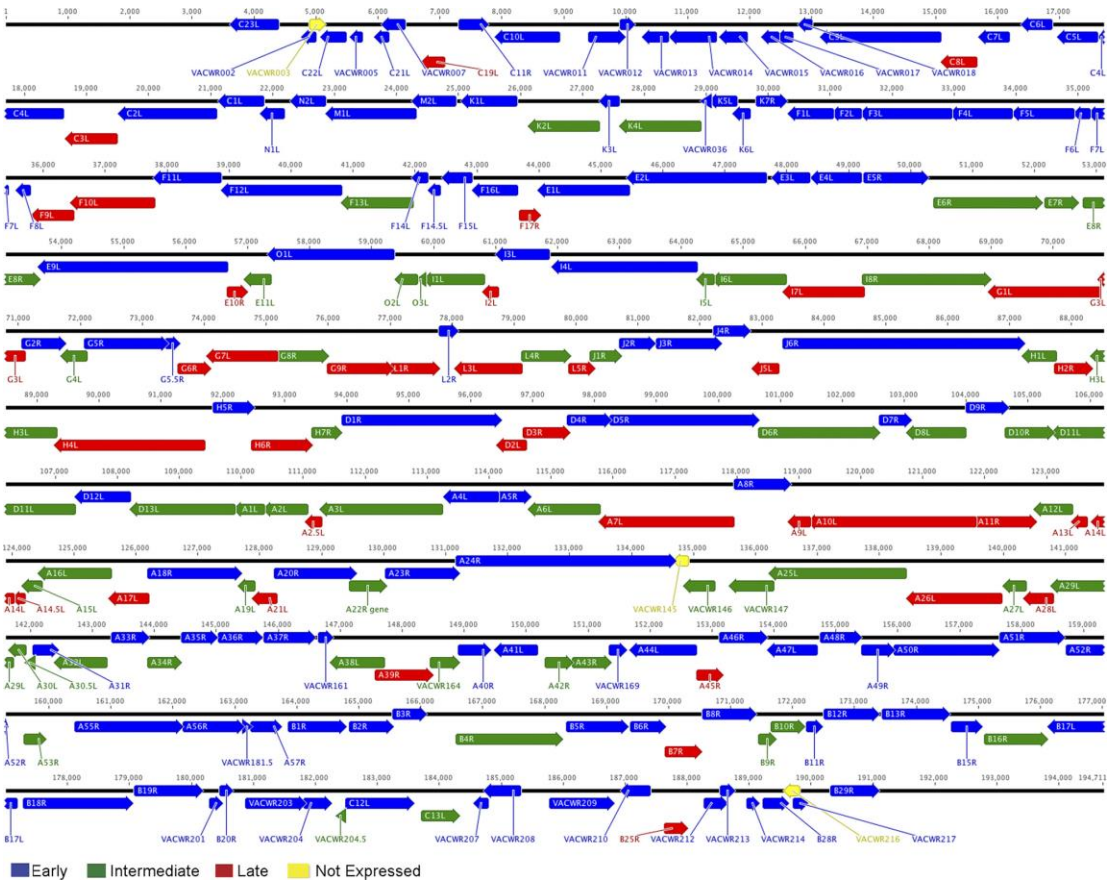

Figure 2.3. VACV transcriptome map. VACV WR ORFs are shown as colored arrows indicating the direction of transcription. When applicable, the common (Copenhagen) HindIII fragment letter/number name was used to identify ORFs; otherwise the VACV WR name was provided. The numbers from 1 to 194711 indicate the nucleotide positions on the VACV genome. Note that each ORF has been assigned the stage at which its earliest expression can be detected and that the presence of additional promoter elements that potentially contribute to later stages of gene expression are not indicated. Figure 8 from Yang, Z., et al. Expression Profiling of the Intermediate and Late Stages of Poxvirus Replication. J Virol, 2011. 85(19): 9899-9908. (3). Reprinted with permission.

2.5.1 Regulation of Early Transcription

As mentioned above, a complete early transcription system is contained within the core of an infectious viral particle. This system produces mRNA that is capped, methylated, and polyadenylated, even *in vitro*. Shortly after entering a cell, the virus core is transported on microtubules to the site of transcription, where mRNA synthesis is detectable within 20 mins. RNA/DNA hybridization studies done with VACV have shown that about half of the genome is transcribed prior to DNA replication, meaning that about half of the genes are early genes (7). These findings have been supported by genome tiling and deep RNA sequencing experiments (40, 97). As mentioned above, these genes are involved in DNA replication and intermediate gene expression, but also include genes that encode for growth factors and immune defense molecules (7). Early transcripts can actually be divided into two groups, based on their kinetic cluster analysis. Both groups are expressed in the presence of protein synthesis inhibitors and are thus classified as immediate early genes (40, 97). The paper, by Yang et al. (97), named the two groups E1.1 and E1.2. There was no distinction between the two regarding their functional class. Both are involved in transcription, DNA replication, and host interactions, and are distributed throughout the genome. One observed difference, however, was in the promoter sequences, as the E1.1 promoter more closely resembled the consensus sequence than E1.2. The authors postulated that the different kinetics may be due to higher-affinity TF binding sites for the E1.1 promoters, other DNA-protein interactions, and DNA packing within the core but the significance of this difference is not yet clear.

Early gene expression terminates upon uncoating, the disruption of the virus core. This latter process is dependent upon RNA and protein synthesis occurring. If protein synthesis inhibitors are used to block uncoating, early mRNA synthesis is both increased and prolonged. In contrast, the use of DNA replication inhibitors does not block uncoating and only prolongs early transcription to a limited extent. These findings suggest that, under normal conditions, disassembly of the core results in the disruption of the early transcription complex, which EM images suggest occurs via the nucleoprotein complex passing through breaches in the core wall (7). As mentioned earlier, the viral type I DNA topoisomerase plays a role in transcription, specifically early transcription. While it is not essential for replication in cell culture, a deletion mutant was found to have a reduction in infectivity of about a log. These mutant particles were able to able to go through all the initial steps of infection but had a severe defect in early gene transcription and a delay in DNA replication. This defect, however, could be overcome by increasing the amount of virus used by 10-fold. These findings were quite surprising, especially given the usual and expected functions of a topoisomerase, i.e., in DNA replication. It seems that the primary and possibly the only role for the viral topoisomerase is to aid in early gene expression (83).

The first step in uncoating begins almost immediately after infection, consisting of the removal of the lipoprotein envelope and liberation of the cores into the cytoplasm. This step obviously requires fusion of the viral envelope or membrane with that of the host plasma or endosomal membrane. The second step of uncoating, which is not as well-characterized as the first step, involves rupture of the core wall

and liberation of the genome, which allows for the subsequent process of DNA replication (73) and signals the end of early transcription. It is this second step which requires RNA and protein synthesis (7, 73). One uncoating protein has been identified, D5, which is also the viral primase/helicase, meaning it is also involved in viral DNA synthesis (73).

Early transcription peaks around 1-1.5 hpi but declines rapidly after that. The decline is so rapid that it cannot solely be explained simply by the cessation of transcription and is consistent with an increase in mRNA degradation after virus infection (7). This enhanced mRNA degradation was at first only speculative (97) but has since been confirmed. It is carried out by the D9 and D10 viral decapping enzymes, leading to the generation of uncapped mRNAs (99, 100), which are presumably rapidly destroyed by host 5` exonuclease activity (40), in the form of Xrn1 (93, 101, 102). This increased rate of mRNA degradation applies to all classes of mRNA (7, 99-102) and it is thought that this rapid turnover serves as a mechanism for eliminating both host and viral mRNA at the end of early and intermediate stages (40, 99, 101). These enzymes would also seem to help emphasize the cutoffs between the early, intermediate, and late phases of gene expression (101). Degradation of host mRNAs would contribute to the observed global host protein shutoff seen in VACV-infected cells, allowing the virus to first have preferential and then exclusive access to translational machinery and macromolecular building blocks (99, 100, 103-105). A paper by Katsafanas and Moss (105) put forth an additional mechanism to explain the global shutdown of host gene expression: sequestration of

the translation initiation factors within the viral factories, where viral transcription and translation take place.

Despite the global host shutoff, a subset of host mRNAs was actually found to be increased at 2 hpi (97). These mRNAs were found to be involved in apoptosis, ligand-mediated signaling, the NF-kB cascade, and signal transduction, likely representing the host response to infection. At 4 hpi, the only increase in cellular mRNAs was seen for those involved in chromatin packaging and assembly (97). Two recent papers explored this topic in more depth (103, 104). The first, by Dai et al. (103), found that host mRNA depletion was predominantly responsible for host protein synthesis shutdown but that despite this, a subset of host mRNAs experienced a significant increase in their translation efficiency. These mRNAs were transcribed from genes involved in oxidative phosphorylation, which generates energy for the cell. This is important because viruses are unable to produce energy on their own and require the cell to survive long enough for them to complete their lifecycle. Their results showed that the short 5` untranslated region (UTR) of the oxidative phosphorylation mRNAs contributed to this observed translational upregulation. Even more telling was that the protein levels and activities of these gene products increased due to VACV infection, in addition to oxygen consumption. Not surprisingly, inhibition of this process greatly reduced viral replication (103). This mechanism contrasts with that of Influenza A Virus (IAV), which protects oxidative phosphorylation simply by maintaining the levels of the appropriate mRNAs and selectively degrading others (104).

The viral decapping enzymes also contribute to virulence (100, 101) directly, by reducing synthesis of immunomodulatory proteins (99) and accelerating host mRNA decay, especially of innate (97, 101) and adaptive immune responses (100), and indirectly, by preventing or reducing the formation and accelerating the degradation of viral double-stranded RNA (dsRNA) (101, 102). dsRNA formation, which is an issue particularly during intermediate and late transcription, is a hallmark pathogen-associated molecular pattern (PAMP) and thus a potent inducer of the host innate antiviral immune response (101, 102, 106). The last benefit of these enzymes is that they eliminate inhibitory capped RNA degradation products or abortive transcripts, which are responsible for translational defects. These molecules can severely hinder viral protein synthesis by competing with viral mRNA for association with ribosomes and/or sequestering translation factors (102).

2.5.2 Early-Stage Promoters and Termination Signal

The promoter for early genes is an A/T-rich region located upstream of the RNA start site. A critical core region was defined from -13 to -27 (15 nucleotides (nt) (107)), relative to the RNA start site, which is denoted as +1. The consensus core sequence was determined to be AAAAAATGAAAAAA/TA, with initiation of transcription occurring with a purine predominantly found 12-17 nt downstream of the core region. This consensus sequence is close to the optimal one, as defined by saturation mutagenesis experiments (7). Interestingly, this core motif is present in several locations throughout the genome and seems to be required but not sufficient, on its own, for the initiation of transcription. Perhaps even more intriguing though is that there appear to be a greater number of transcription start sites (TSSs) versus

annotated ORFs. This brings up the possibility that the virus may have the capacity to generate new proteins over the course of its evolution (40, 107). Most of these putative or anomalous TSSs lack an upstream core motif, however. Most mapped to the positive-sense ORFs. Those that were found associated with a highly conserved core promoter motif (about 5%), though, were capable of being translated into shorter protein isoforms (107). The DNA that lies between the core and the RNA start site appears to play a spacer role (7) and is 10 nt in length. This region may facilitate DNA unwinding, which would form the open complex needed for transcription initiation (107). Additional studies have revealed that a high AT content in this spacer region is correlated with active transcription (107). There is also no evidence suggesting the presence of transcriptional enhancer elements (7, 97), as has been found in nuclear DNA viruses (7). The distance between the 5` boundaries of early mRNA and the first AUG codon varies considerably but has a mean of about 40 nt. Due to the lack of enhancer elements, it was hypothesized that perhaps both the level and timing of early gene expression is regulated by the promoter sequence (97). Promoter regions are conserved between different POXV genera (36-38), which provided an explanation for an old phenomenon referred to as non-genetic reactivation. In this scenario, a heat-inactivated POXV is rescued via coinfection with another POXV from a different genus, with the heat-killed virus providing the template while the other POXV provides the enzymes needed for transcription (7).

VACV early mRNA transcription terminates between 20 and 50 bp downstream of the RNA sequence UUUUUNU (TTTTTNT in the DNA). This termination sequence has been found at the ends of most, but not all, early genes.

When it is not present, the mRNA tail may extend over into the next downstream early gene (40). Experimental evidence suggests that the termination efficiency is about 80%, although it is less in some cases due to the RNA secondary structure (7). This 80% termination efficiency applies to genes with a single termination motif, which most early genes have. Some, however, have multiple termination signals, while others have none. Still, in others, the signal is unrecognized because of the secondary structure. As such, the RNAP may continue to transcribe downstream ORFs until encountering functional termination sequences. This scenario would potentially lead to gene misclassification, as mentioned above. If the two adjacent genes were both early, read-through would not affect gene classification. If an early gene, however, overlapped with an adjacent intermediate or late gene, then the latter would be wrongly categorized as early (97). This termination signal has also been found near the ends of putative early genes in other POXV genera, suggesting it plays a similar role in termination in other POXVs besides VACV (7). A genome-wide analysis of the 3' ends of VACV early mRNAs revealed that only about two-thirds actually contained the termination sequence, raising the possibility that an additional termination mechanism may exist. The mechanism in question may have been revealed when the existence of a pyrimidine-rich sequence in the coding strand up to position -25, relative to the polyadenylation site (PAS), was found. This sequence was found regardless of the termination sequence being present, suggesting that it may facilitate termination (40, 107). Specifically, this mechanism may involve the viral protein NPH II. This protein has RNA helicase activity and efficiently unwinds DNA-RNA hybrids containing a purine-rich DNA tracking strand. Also, of note is

the fact that virions lacking NPH II synthesize RNA molecules that are abnormally long and inefficiently released from the viral core. Thus, the observed enrichment of pyrimidines in the coding strand, which corresponds to purines in the tracking strand, upstream of the PAS may facilitate termination. The first PAS was usually around 40 nt after the termination sequence for most early transcripts. Additional PASs can be found downstream of the initial one. These other ones may provide backup termination or serve to terminate anomalous transcripts. They may even represent re-polyadenylation of processed RNAs (107).

2.5.3 Enzymes and Factors for Early-Stage Transcription

Many virally encoded enzymes and factors are packaged into the virus particle, as evidence by the fact that a soluble extract of VACV virions can transcribe an early promoter template *in vitro* and produce properly initiated and terminated mRNA. Some of these packaged enzymes and factors include the multisubunit DNA-dependent RNA polymerase (MSDdRP), RAP94, VACV early transcription factor (VETF), capping and methylating enzymes, the poly(A) polymerase, NPH I, and topoisomerase, all of which are directly involved in the synthesis and/or modification of mRNA. Additional packaged enzymes have various other roles in processes such as virion assembly, protein processing, and DNA packaging (82). The vRNAP resembles those found in eukaryotes regarding their size and complexity. It is encoded by at least eight viral genes, and the subunits range in size from 7 to 147 kDa. While it is known that at least one copy of each large subunit is present in the complex, the stoichiometry of the small subunits is still unknown (7), although if each subunit is only present once, including RAP94, it would have a mass of at least 500

kDa (108). Multiple subunits resemble their counterparts in eukaryotic cells. For instance, RNA polymerase subunit (RPO) 147 (J6) and RPO132 (A24) have 20-30% sequence identity with the corresponding large subunits of cellular RNA polymerases, RPO30 (E4) is about 23% identical to eukaryotic transcription factor SII (TFIIS), and RPO7 is homologous (about 23% amino acid sequence identity) to the smallest eukaryotic RNA polymerase subunit. In fact, all of the subunits, with the exceptions of RPO30 and RPO35 (A29), are homologous to cellular RNA polymerases, suggesting they perform similar duties for the virus (40).

RAP94, encoded by the H4L gene, is associated with about half of the RNAP molecules in VACV virions. This protein is specifically required for transcribing from early promoter templates, along with VETF. RNAP is still able to transcribe without RAP94, but that transcriptional activity is limited to single-stranded DNA nonspecifically or double-stranded intermediate and late promoter templates with the appropriate transcription factors. RAP94 is made only during the late phase of expression, which aligns well with its exclusive role as a virion-associated early TF. This aspect sets it apart from the other core RNAP subunits, which all have early promoters, as they are needed for intermediate and late transcription (40). It seems likely though that de novo synthesis of RNAP occurs prior to both intermediate and late gene transcription, especially given that RNAP in complex with RAP94 is specific for early gene transcription (108).

VETF, as with RAP94, is made only during the late phase of expression. VETF is a heterodimer, formed by the 82- and 70-kDa proteins expressed by the A7L and D6R genes, respectively. VETF alters the DNA conformation as a result of its

dual binding sites, one in the core region of early promoters and the other in DNA downstream of the RNA start site (40, 109). This downstream region of DNA need not have a specific sequence, as it seems to simply aid in stabilizing the transcriptional complex. This dual interaction seems to be the responsibility of only one molecule of VETF (109). Mutations in the core region that decrease transcription also negatively affect VETF's ability to bind the DNA, leading to the conclusion that it binds this stretch of DNA in a sequence-specific manner (40, 109). It was also determined that the small subunit of VETF possesses DNA-dependent ATPase activity that is essential for transcription, but not for promoter binding. The mechanism for this essential function may be via promoter clearance. RNAP, RAP94, and VETF are sufficient for transcriptional activity *in vitro*. Complexes of VETF and RNAP have been found, raising the possibility that VETF recruits RNAP to the promoter. It also seemed likely that there was a direct interaction between VETF and RAP94, a theory that has since born out to be true (40, 110). This study by Yang and Moss (110) yielded many findings: VETF associates only with RPO containing RAP94 (both *in vitro* and *in vivo*); the association between RAP94 and VETF requires both subunits of VETF; neither viral DNA nor other virus-encoded late proteins are required for the interaction of RAP94 with VETF and core RPO subunits; different domains of RAP94 bind VETF and core subunits of RPO; and VETF and NPH I bind independently, and possibly simultaneously, to the N-terminus of RAP94. As such, they concluded that RAP94 serves as the bridge connecting RPO and the proteins necessary for transcription initiation, elongation, and termination.

Once transcription has been initiated, the process of elongation of the transcript occurs. The elongation complex possesses 3` RNase activity which allows a stalled polymerase to resume transcribing. It has been noted that such activity resembles that exhibited by eukaryotic RNA polymerase II in the presence of TFIIS, which the VACV RPO30 subunit happens to be a homolog of (7). A complex consisting of RNAP, RAP94, VETF, capping enzyme (alternatively referred to as vaccinia termination factor or VTF (111) with regard to its role in termination (40)), and NPH I is able to both initiate and accurately terminate transcription on DNA templates containing an early promoter and the TTTTTNT termination signal. While RNAP and VETF are sufficient for reconstitution of transcriptional initiation and elongation, capping enzyme and NPH I, a DNA-dependent ATPase, are required for the release of nascent mRNA with the UUUUUNU termination sequence from the transcriptional complex. This transcript release process requires ATP. It has also been demonstrated that NPH I interacts with RAP94, which may provide an explanation for the specificity of this termination complex for early transcripts (7). NPH I is capable of releasing transcripts independent of VTF and the termination sequences, as long as the vRNAP is not actively undergoing transcription elongation. If the vRNAP is actively elongating, however, then both VTF and the termination sequence are required (111). A recent paper (111) found that VTF, in the presence of the termination signal, enhances pausing of the vRNAP, thus inhibiting elongation and facilitating NPH I-mediated transcript release. The authors of this paper also put forth an alternative hypothesis with regard to pyrimidine-rich sequences possibly functioning to mediate transcript release (107). They thought that the secondary

structures of most early mRNAs alone may not be enough to induce transcriptional pausing and termination. These transcripts would then require VTF and the termination signal. For those few that do not contain the termination signal, however, the RNA secondary structure may be sufficient to cause termination on their own, with VTF functioning to stabilize these weak RNA structures (111). An additional role NPH I plays is that of an elongation factor for the polymerase to facilitate read-through of intrinsic pause sites (7).

Early transcripts, whether made *in vitro* or *in vivo*, are capped and polyadenylated such that they resemble, structurally and functionally, eukaryotic mRNA. These transcripts possess a cap I structure, which consists of a terminal 7-methylguanosine connected to a 2`-O-methylribonucleoside via a triphosphate bridge. While the role of the ribose methylation has yet to be determined (112), the N^7-methylguanosine portion of the cap is essential for the stability of the mRNA and for the binding of the viral mRNA to ribosomes. Capping takes place co-transcriptionally, when the nascent RNA transcripts reach about 30 nt in length. The capping enzyme is a heterodimer, consisting of the D1R and D12L gene products, and it carries out the first three steps of cap formation. The 97-kDa (D1R) large subunit forms a covalent lysyl-GMP intermediate in addition to possessing both the RNA triphosphatase and guanylyltransferase activities in its N-terminal region. The N^7-methyltransferase activity, on the other hand, is formed by the C-terminal region of the large subunit and the 33-kDa small subunit. Cap formation occurs as follows: first, the removal of the terminal phosphate of the triphosphate end of the nascent RNA to form a pp(5`)N-terminus; next, transfer of a GMP residue from GTP to form

G(5`)ppp(5`)N-; then, transfer of a methyl group from S-adenosylmethionine to produce m7G(5`)ppp(5`)N-; and finally, transfer of a second methyl group to form m7G(5`)ppp(5`)Nm. This final step is actually carried out by a separate viral protein, VP39, which is the gene product of J3R. Viral mRNAs made *in vivo* possess additional base and ribose methylations catalyzed by cellular enzymes (7). Phenotypic analysis studies carried out with a ts mutant of the capping enzyme large subunit support its multifunctional roles (40).

Interestingly, VP39, which is a nucleoside 2`-methyltransferase, exists in two forms: as a monomer and as the small subunit of the poly(A) polymerase, which itself exists as a heterodimer, consisting of VP39 and VP55 (E1L). VP55 is responsible for binding to uridylate sequences near the end of the RNA transcript and catalyzing the processive addition of 30-35 adenylate residues before switching over to a slow, non-processive mechanism. VP39's role as part of this heterodimer is to bind poly(A) and stimulate VP55 to semi-processively add additional adenylate residues. Thanks to its dual function as a methyltransferase and as a processivity factor for the poly(A) polymerase, VP39 is expressed in excess, compared to VP55, in infected cells. Interestingly enough, the two functions of VP39 are independent of each other, as studies found that mutations in VP39 that abolished its methyltransferase activity had no effect on its adenyltransferase stimulatory activity. The revelation that the capping enzyme/termination factor activities and the ribose methyltransferase/poly(A) polymerase processivity factor activities are housed within the same enzymes is quite fascinating. These seemingly disparate functions operate at opposite ends of the mRNA, and it is interesting to speculate whether these

peculiar pairings may provide a specific advantage to the virus or simply represent an economical use of its proteins. Genetic and biochemical studies had implicated VP39 playing a role in intermediate and late transcription elongation (7), which has since been confirmed (72).

The minimal complex for synthesis of correctly initiated, capped, terminated, and polyadenylated early mRNAs was determined using *in vitro* reconstitution assays. There are, however, additional proteins required within the viral core. They are: the H6 DNA topoisomerase I, the I8 NPH II, and H1 serine/tyrosine phosphatase, and L4 DNA-RNA binding protein, and the L3 and E8 proteins, which have no known functions as of yet (7, 40).

2.5.4 Regulation of Intermediate Transcription

Intermediate gene expression cannot take place until DNA replication occurs. This requirement may stem from the fact that the genome is inaccessible to the newly synthesized transcription factors prior to uncoating and replication. This potential inaccessibility may be due to specific repressor proteins, which is less likely, or nonspecific due to remaining virion proteins. This inaccessibility is supported by transfection experiments demonstrating that DNA isolated from purified virions can serve as a template for intermediate and late transcription without DNA replication occurring (7). Additional support for this idea is the fact that a transfected intermediate gene can get transcribed prior to DNA replicating taking place, even in the presence of a DNA synthesis inhibitor like hydroxyurea, whereas that same gene in the viral genome does not get transcribed under the same conditions (113, 114). In fact, the levels of transcription seen are higher than when DNA replication is allowed

to proceed normally (108). Intermediate gene expression can be detected around 100 mins after infection (3), peaks around 2 hpi, but rapidly declines from there, similar to the decline observed for early gene expression. The rapid decline seen for intermediate mRNA, as with early transcription, is a combination of two factors: termination of transcription and rapid mRNA turnover. Initially, only five VACV intermediate genes had been characterized: the late transcription factors A1, A2, and G8 (7), (115), the I8 RNA helicase NPH II (7), and the I3 single-stranded DNA binding protein, which interacts with ribonucleotide reductase and is regulated by both early and intermediate promoters (7, 113). Since then, the list of intermediate genes has been expanded to 53 (3, 40), although some of these are expressed late as well (3). Gel electrophoresis analysis of intermediate mRNA shows diffuse bands equal to and longer than the coding regions, suggesting that the preferred sites of 3`-end formation that do not correlate with the TTTTTNT, the early transcription termination sequence (7).

2.5.5 Intermediate-Stage Promoters

The core region of intermediate promoters resembles that of early promoters in terms of A/T richness, but they differ with regard to the specific sequence. Another similarity between the two is that there is no evidence of enhancer elements. Intermediate promoters possess two important regions, as determined by mutagenesis experiments: a 14-bp core element separated by 10 or 11 bp from a 4-bp initiator element, TAAA. Intermediate-stage mRNAs are initiated within the AAA, but these areas have incorporated additional A residues thanks to a polymerase slippage

mechanism (7). A consensus sequence was derived from analysis looking at all the sequences preceding the 53 intermediate ORFs that have been identified (40).

2.5.6 Enzymes and Factors for Intermediate-Stage Transcription

Both viral and cellular proteins are involved in intermediate transcription as intermediate promoter templates can be transcribed using VACV-infected cell extracts in the presence of a DNA replication inhibitor. The viral proteins involved include: the vRNAP; the viral capping enzyme, which in addition to capping the mRNAs (7) acts as a transcription initiation factor (116); VACV intermediate transcription factor (VITF) -1, which is RPO30; and VITF-3, a heterodimer formed by the 45-kDa A23 and 34-kDa A8 proteins (7, 117). A8 and A23 are essential for the virus. Without them, intermediate and late gene expression and resolution of concatemers do not occur, but early gene expression and DNA replication still take place. As a result, deletion mutants must be studied using a complementing cell line (118). It has also been reported that protein phosphorylation is involved in intermediate transcription, via the B1 kinase. B1 mutants were found to be defective in intermediate transcription but not late transcription. The substrate(s) that may require phosphorylation to function during this process have not been determined (108). Unlike VETF, neither VITF-1 nor VITF-3 exhibit ATPase activities or sequence-specific DNA binding (7). Both VITFs have early promoters (7) and are thus synthesized during early protein synthesis, prior to DNA replication (113). This means they can begin aiding in transcribing intermediate stage genes once the viral DNA has been sufficiently copied. VITF-2 was determined to come from the host, as it was originally extracted from the nucleus of uninfected HeLa cells but was found to

be distributed throughout the nuclear and cytoplasmic fractions of VACV-infected cells. Two different proteins, which form a complex together, were found to serve the role of VITF-2 in an *in vitro* complementation assay: G3BP, the Ras-GTPase-activating protein SH3 domain-binding protein, and p137/Caprin-1, the cytoplasmic activation/proliferation-associated protein (7, 119). These proteins have been observed to accumulate at sites of transcription in viral factories, though an essential role *in vivo* has yet to be confirmed (105). One interesting theory for the role of VITF-2 is that it may act as a gatekeeper between the pre- and post-replicative phases of the virus lifecycle by signaling that a quiescent cell has been activated, which would allow for optimal viral replication (7, 119). Additional studies have revealed that the cellular TATA-binding protein positively affects VACV intermediate and late gene expression, while the eukaryotic transcription factor YY1 negatively affects post-replicative gene expression (40). In the case of the latter, it was found that YY1 binds, *in vitro*, to the promoter of the I1L gene, which was initially thought to possess a late promoter but has since been categorized as having an intermediate promoter. Searching through the VACV genome revealed numerous other promoters containing the YY1 binding site, TAAATGG. Furthermore, YY1 was found to localize to the cytoplasm of infected cells (7). Such a reliance on the host is a bit surprising given that POXVs encode their own RNA polymerases and other transcription factors (40).

Multiple genetic studies have supported the idea that positive and negative regulators of intermediate and/or late transcription exist. One example involves VACV A18 RNA helicase mutants, which have an indistinguishable phenotype, with regard to RNA metabolism, from wild-type (wt) VACV infection in the presence of

the antiviral drug isatin-B-thiosemicarbazone (IBT) (7, 120). Specifically, what is observed is an increase in the steady state concentrations of late mRNA, activation of the cellular ribonucleolytic 2-5A pathway, which results in cleavage of viral mRNA and host rRNA and premature cessation of late protein synthesis (120). Mutations which conferred resistance to the drug mapped to RPO132 (A24R) (7, 120, 121), while drug-dependent mutations mapped to the G2 protein (7, 120), which has since been determined to be a positive transcription elongation factor (72). The IBT resistant mutant that was isolated (120) did have to endure a fitness trade-off though, as it was discovered that this mutant was defective in transcription elongation, with the mutant complex pausing longer and more frequently than the wt complex (121). Since then, additional drug-resistant mutants have been mapped to RPO147 (J6R) (122). Adding another wrinkle to this story is the fact that while mutations in A18 resulted in increased transcriptional read-through, mutations in G2 resulted in decreased mRNA length. Furthermore, G2 mutants can rescue A18 mutants. Physical interactions have been demonstrated to take place between G2, A18, and H5, one of the late transcription factors (7, 123). A18 has also been implicated as a transcript release factor, working in conjunction with a yet unidentified cellular protein. It was also determined that J3R also had a role as a transcription elongation factor, in addition to serving as the cap 2`-O-methyltransferase and as a subunit of the poly(A) polymerase (7). Additional IBT-dependent mutants have been mapped to the J3R gene. These mutants, along with the drug-dependent mutants mapped to G2R, produce intermediate and late transcripts, *in vivo*, which are shorter than wt transcripts (122).

2.5.7 Regulation of Late Transcription

Late gene transcription can be detected around 140 mins after infection (3), peaks around 4 hpi, and continues for the remainder of the viral lifecycle. The need for such continuous mRNA synthesis is emphasized by the fact that late mRNAs have an estimated half-life of about 30 mins or less. This high level of transcription leads to the accumulation of large amounts of late proteins, which include VETF, RAP94, and the major virion components (7). The late genes, although spread through the genome, tend to cluster in the central region. A study which was undertaken to distinguish between intermediate and late genes identified 38 total that belonged to the latter phase of gene expression (3, 40).

2.5.8 Late-Stage Promoters

The late promoter can be broken down into three regions: a core sequence of about 20 bp with some consecutive T or A residues, which is separated from a highly conserved TAAAT element in which transcription initiates, by a 6-bp region. Mutations of the TAAAT region severely decrease transcription, while a synthetic promoter consisting entirely of Ts in the core region proved stronger than any natural late promoter that was tested (7). The consensus promoter sequence that was derived from analysis of 38 late ORF upstream sequences corroborated the conclusions from the mutational analysis (40). This region is usually followed immediately by a G and sometimes an A. The former scenario results in a transcription initiation sequence of TAAATG, which results in an overlap with the translation initiation codon, ATG. The apparent lack of an untranslated RNA leader in this situation was a source of confusion, until it was discovered that late mRNAs contain capped, heterogeneous-

length poly(A) leader sequences formed from slippage of the RNAP. The poly(A) leaders have also been observed on certain early mRNAs with a TAAAT initiation site and on intermediate mRNA. These findings seem to indicate that slippage on the complement of the AAA sequence is an intrinsic property of the vRNAP (7).

2.5.9 5` Poly(A) Leaders of Post-Replicative Genes

One possible benefit of the structure and nature of post-replicative genes is that the 5` poly(A) leader could compensate for the complementary RNA by providing a single-stranded binding site for the 40S ribosomal subunit and initiation factors, which would then move by antisense RNA, unimpeded, to the first AUG, where ribosome assembly and translation would take place (7). The 5` poly(A) leaders, which are present on all post-replicative mRNAs but only a few early mRNAs, may reduce the need for certain translation initiation factors (102). These 5` poly(A) leaders may also facilitate translation and stability (40). These hypotheses were put to the test in a recent paper (124). What this group found was that the 5` poly(A) leaders of post-replicative genes did indeed confer a translational advantage during infection. While a constitutive and uninterrupted leader 12 residues long was optimal, the most frequent lengths were between 8 and 12 residues, supporting findings from another study (125). The consistency of the length of the leader led the authors to conclude that this process has been optimized over the course of POXV evolution in order to maximize protein synthesis, and not simply an accident due to the vRNAP (5, 124). This advantage was even seen in infected cells with impaired cap-dependent translation (124), but not in uninfected cells (5).

Furthermore, this paper raised the possibility that the 5` poly(A) leaders may allow for cap-independent translation, although it does not seem to serve as an IRES. Such a requirement would be necessary, since intermediate and late genes are made after host protein synthesis has been shut down. This shutdown would include cap-dependent translation initiation factors. Additionally, cap-independent translation would ensure efficient translation of viral mRNA under other physiological conditions, like cell mitosis and cellular stress, during which cap-dependent translation is suppressed. This would allow VACV to switch between the two forms of translation depending on the availability of translation initiation factors and the status of the cellular environment, which has also been observed to happen with cellular mRNAs (124). This capacity for utilizing two modes of initiation probably maximizes how well viral mRNAs compete for ribosomes or allows them to undergo initiation in spite of the viral decapping enzymes, which do their job indiscriminately (**Fig 2.4**).

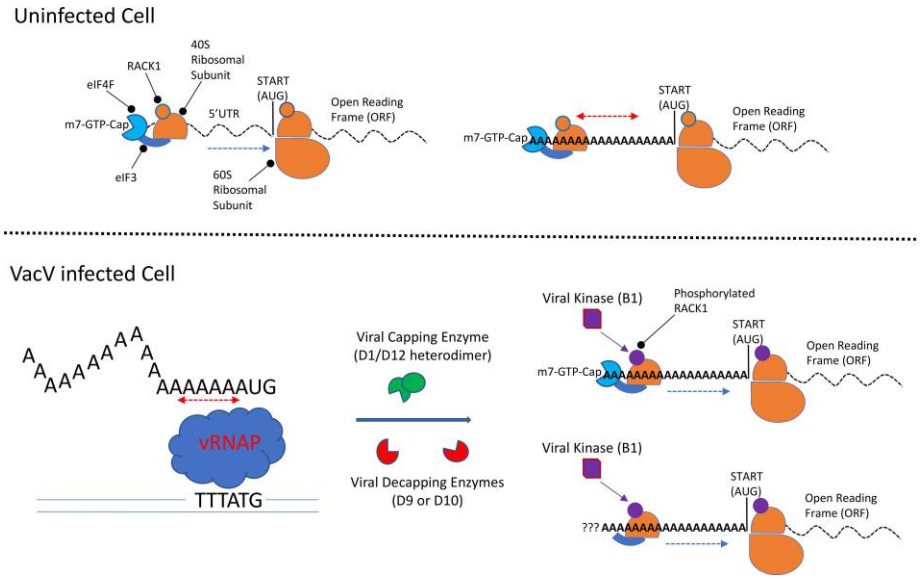

Figure 2.4. Translation control in uninfected and VACV-infected cells. In uninfected cells (top), the 40S ribosomal subunit is recruited to the mRNA m-7-GTP cap through the combined actions of eIF4F and eIF3. The ribosome then scans (blue arrow) the 5` UTR in search of a start codon, at which point the 60S ribosomal subunit joins to initiate translation. Ribosomes slide (red arrows) on polyA stretches, and 50 polyA leaders do not act as enhancers in mammalian cells. In VacV-infected cells (bottom), the vRNAP slips on intermediate and late promoters, reiterating adenosine residues to create randomly sized polyA leaders. Viral enzymes control mRNA capping and decapping in infected cells, where polyA leaders enable either cap-dependent or cap-independent translation of viral mRNAs; whether cap-independent translation occurs on viral mRNAs that have been decapped or were never capped remains unknown (?). Modification of RACK1 by the viral B1 kinase contributes to the ability of polyA leaders to function in infected cells. eIF, eukaryotic initiation factor; m-7-GTP, 7-Methylguanosine-5`-triphosphate; RACK1, receptor for activated C kinase 1; UTR, untranslated region; VacV, vaccinia virus; vRNAP, viral RNA polymerase. Adapted from Figure 1 from Walsh, D. Poxviruses: Slipping and sliding through transcription and translation. PLoS Pathog, 2017. 13(11): e1006634. (5). Reprinted with permission under the Creative Commons Attribution License.

It is also possible that the high abundance of viral mRNAs contributes somewhat to the apparent eIF4F-independent phenotype (5). The group, however, did not rule out the possibility of a minimal amount of cap-binding translation initiation factor eIF4E being required, which would be an alternative cap-dependent translation mechanism (124). It had already been shown that VACV sequesters crucial translation initiation factors within the viral factories (105). It should be noted, however, that once a ribosome has been recruited to a viral transcript, these 5` poly(A) leaders may become a liability. Runs of 11 or more As cause bidirectional sliding of ribosomes in a process known as phase-less wandering. Adding another wrinkle to this story is the recent discovery that the B1 kinase acts upon residues in a flexible loop in the small ribosomal subunit protein, receptor for activated C kinase 1 (RACK1). This phosphorylation appears to slow initiation rates to facilitate leader activity, which may compensate for sliding and/or allowing cap-independent viral transcripts to dominate (**Fig 2.4**). Thus, VACV relies on slippage during both transcription and translation, in a peculiar coupled strategy (5). Dhungel et al. also found that the 5` poly(A) leader could be used to increase foreign gene expression in conjunction with POXV-based vectors (124), such as the bacteriophage T7 promoter-based expression system with VACV (27, 126, 127) that was used in this study. Taking a broader view, this paper also brings up the possibility that their findings could point to a novel cellular translational mechanism (124).

2.5.10 RNA Processing Signal

As with intermediate transcription, the early transcription termination signal is ignored during late transcription, resulting in TTTTTNT frequently being a part of the

coding sequence. Because of this, most late mRNAs are long, heterogeneous, and lack defined 3`-ends. This terminal heterogeneity, coupled with concurrent transcription from both strands of DNA, commonly causes overlapping of transcripts. This leads to self-annealing of the late transcripts or annealing with early transcripts, which result in ribonuclease-resistant hybrids (7). In fact, at 4 hpi, observations have been made of RNA sequences covering both DNA strands in their entirety, which is consistent with the annealing of late RNA forming long duplex structures (97). These structures may be resolved or the deleterious effects may be inhibited by the virally-encoded dsRNA binding protein, E3 (101, 102), the K3 protein, the K1 protein, (101), the D9 and D10 decapping enzymes (101, 102), or the RNA helicase (7). Possessing many ways to deal with the production of dsRNA is important since it activates the host's innate immune response. dsRNA is made during the replication of most RNA and DNA viruses (101), making it a signature feature of viral infection,. In the case of VACV, about 15% of the polyadenylated RNA made during intermediate and late gene expression can anneal, forming long intermolecular duplexes with single-stranded RNA tails (101, 102).

There are at least two known exceptions to the general 3` heterogeneity of late mRNAs. In CPXV, the late mRNA encoding the ATI protein has a 3`-end corresponding to a specific site in the DNA template. The DNA sequence at this site encodes an RNA cis-acting signal for RNA 3`-end formation, which functions independently of either the nature of the promoter or the RNAP responsible for generating the primary RNA. Evidence supports the idea that a specific endonuclease is induced or activated to cleave this RNA, which is then polyadenylated. The F17R

transcript of VACV is generated via cleavage as well (7). Endonuclease activity attributed to the H5R gene product is capable of cleaving the F17 RNA, which is then polyadenylated (40). It is unclear how many other late mRNAs are processed similarly. This means that POXVs have two different mechanisms for generating the 3`-ends of RNA transcripts: a sequence-specific manner, in the form of an RNA signal, which is used during the early phase of infection; and a site-specific manner, involving cleavage, used during the late phase of infection (7).

2.5.11 Enzymes and Factors for Late-Stage Transcription

There are four confirmed VACV late transcription factors (VLTFs): G8, A1, A2, and H5, which are referred to as VLTF-1, -2, -3, and -4, respectively. The first three all have intermediate promoters (7), meaning they are only transcribed from replicated or transfected viral DNA. These three TFs were also found to be both necessary and sufficient to overcome a block in late promoter expression (115). H5, on the other hand, has an early promoter (7). A1, A2, and G8 are all essential for viral replication, so deletion mutants must be characterized using a complementing cell line (98), similar to the VITF-3 subunits A8 and A23 (118). A1 and A2 can both bind zinc (7, 115), while yeast two-hybrid (Y2H) studies provide evidence of interactions between G8 and A1 and G8 with itself (7, 128). Additional interactions were seen A1 and itself and G8 and A2 (128). Computational analysis has suggested that G8 possesses a PCNA-like sliding clamp motif (40). Additionally, other studies have reported interactions between H5 and the other late transcription factors (7, 128), specifically A2 and G8 (128). H5 may also play a role in transcript elongation (40), possibly as part of a higher-order transcriptional complex including G2, A18,

the vRNAP, and other additional transcription factors (123). H5 reportedly also interacts with itself (128). *In vitro* transcription assays have demonstrated that H5 increases late transcription several-fold, while also hinting at the involvement of a host factor, dubbed VLTF-X, consisting of the heterogeneous nuclear ribonucleoproteins (hnRNPs) A2/B1 and RBM3 (7), both of which were found to bind to A2. Additionally, hnRNP A2 was observed to interact with itself (128). Similar to intermediate transcription, evidence suggests that the cellular TATA-binding protein and YY1 play positive and negative regulatory roles, respectively, during late transcription (40).

2.5.12 Intermediate and Late Gene Comparison

Initial attempts to distinguish between the two classes of post-replicative genes were thwarted for a variety of reasons. The promoters are similar, so it was difficult to identify them based on sequence. Even techniques such as RNA-seq and genomic tiling were unable to make distinctions between intermediate and late genes. This was due to the close spacing of the ORFs themselves, the extensive read-through of transcripts, and the small differences in temporal expression levels. However, genes regulated by intermediate and late promoters can be distinguished by transfecting plasmids into cells in the presence of a DNA replication inhibitor, as only the intermediate genes will be expressed (3). One study, by Yang et al. (98), found that, of the 53 intermediate promoters, 26 had detectable late activity. This study used a cutoff of 5% activity in relation to the strong F17R promoter. This paper also found that there were two statistically significant features for late activity, neither of which were correlated with intermediate transcription activity: a T following TAAA

at +4 in the initiator region and T residues between -12 and -8 in the core region of the coding DNA strand. Meanwhile, the frequency of A residues between -22 and -14 was correlated with intermediate transcription activity, but not late activity. The importance of these regions was demonstrated via transformation of a strict intermediate promoter into a late promoter, and vice versa. It had also been noted by a previous study that interconversion of intermediate and late promoters with TAAAT initiator sequences could be accomplished, at least to some extent, by altering the space between the core and initiator sequences (98).

2.5.13 Post-Replicative Gene Initiation and 3` End Formation

Similar to what was found for early genes (107), pervasive initiation and 3` end formation have been observed for both intermediate and late mRNAs. Anomalous TSSs were found were located throughout the genome, within both the coding and non-coding strands. Again, like the early genes, some were downstream of functional promoter sequences and can generate truncated isoforms of proteins. The 5` poly(A) leaders for intermediate and late genes were found to be at least 5 nt long for the vast majority of transcripts. The median lengths were 8 and 11 nt for intermediate and late transcripts, respectively. The poly(A) leaders for anomalous TSSs were shorter in general, perhaps hinting at the fact that they might not be optimal for translation. Interestingly, pyrimidine-rich sequences were found immediately upstream of the PASs, suggesting that NPH-II-mediated termination may occur for both early and post-replicative transcripts (125). Perhaps this reliance on pyrimidine-rich sequences is related to the Katsafanas et al. (77) study in which pyrimidine nucleotide depletion greatly inhibited VACV replication. While the effect

on the DNA synthesis and processing was mentioned earlier, it was also found that transcription and translation were affected. The levels of intermediate and late transcripts were drastically reduced, while the effect on early transcripts was less cut and dry. For at least one example of each, shorter early transcripts (500 nt) saw a two-fold increase, while longer early transcripts (4300 nt) saw a two- to three-fold decrease. The authors postulated that longer mRNAs, or at least ones with long pyrimidine tracts, may be more affected than the shorter mRNAs. They also observed prolonged synthesis of some early proteins and a severe reduction in late proteins.

2.5.14 Viral DNA-Associated Transcription Proteins

VACV encodes an impressive complement of transcription proteins, including an MSDdRP, stage-specific TFs, a type I topoisomerase, an elongation factor, capping and methylating enzymes, a poly(A) polymerase, and transcription termination factors. Despite this repertoire, there are questions as to whether additional proteins are involved in these steps of the viral lifecycle (72). While the notion that POXVs perform functions seemingly independent of their hosts has prevailed for years, it has become increasingly obvious that host proteins are needed for both intermediate and late transcription, in contrast to early transcription. The rationale behind this would seem to be two-fold: location and timing. Early transcription occurs entirely within the confines of the viral core, whereas intermediate and late transcription occur in the cytoplasm, which is a much more open environment. The timing of intermediate and late transcription, after DNA replication, may also account for the recruitment of host proteins. It would seem to

be more beneficial for the virus to postpone recruitment of host proteins to the viral factories until sufficient amounts of DNA have accumulated in the cytoplasm (108). For instance, as mentioned above, roles for host proteins have been observed, including cellular heterogeneous nuclear ribonucleoproteins A2/B1, G3BP and p137 (or caprin I), TATA binding protein, and YY1 (7).

To provide some answers to these open questions, Senkevich et al. (72) performed experiments to try and confirm known and perhaps identify new members of the VACV transcription machinery. They were able to identify viral proteins known to have roles in intermediate and late transcription, which indicated that the pool of replicating DNA was being transcribed. The four largest subunits of the vRNAP (RPO 147, 132, 35, and 30) were detected, but the remaining four were not. Both subunits of the capping enzyme (D1 and D12) and cap 2` O-methyltransferase (J3) were also recovered, in addition to a positive transcription elongation factor, G2. As mentioned before, J3 is also the noncatalytic/small subunit of the poly(A) polymerase and a positive transcription elongation factor. Perhaps the most intriguing result, however, was that the intermediate and late transcription factor complexes were both recovered. This was interesting because none have been shown to directly interact with either DNA or the RNAP. H5, which plays, a role in DNA replication, is also involved in transcription. It also interacts with G2. Host proteins previously reported to be involved in VACV transcription were also detected. These included G3BP1, Caprin-1, and HNRPA2B1. G3BP1 and Caprin-1 form a heterodimer have been demonstrated to co-localize with viral mRNA in virus factories and were isolated from cell extracts while looking for factors needed for

intermediate transcription (72, 105). The RNA-binding proteins HNRPA2B1 and RBM3 have been found to stimulate late gene transcription, as mentioned previously, and the former was detected in this study as being associated with nascent DNA (72).

The large/catalytic subunit (E1) of the poly(A) polymerase was not detected in this study, consistent with its posttranscriptional activity and its association with mRNA instead of DNA. Additional host proteins with suggested roles in transcription that were not found associated with the nascent DNA include RBM3, the TATA binding protein, and YY1. The TATA binding protein and YY1, as mentioned above, reportedly upregulate and downregulate intermediate and late transcription, respectively, via promoter binding, so not detecting them casts some doubt on their biological relevance (72).

2.5.15 Translation and Annotation of Viral ORFs and Transcripts

One consistent theme for POXV gene expression is that proteins which interact with each other are frequently made at the same time. One such example is the partnering of the EFC and the redox proteins. Most members of the EFC have intramolecular disulfide bonds, which are formed by the viral redox proteins. Both sets of proteins are made late in infection. In contrast, the proteins used for the MV membrane and the EV membrane are made at different times, supporting the idea of independent pathways of membrane morphogenesis. It is conceivable that synthesizing some proteins at the wrong time could have a deleterious effect on the virus (3).

As mentioned previously, pervasive initiation and 3`-end formation have been found for both early and post-replicative transcripts. In the cases of both sets of

genes, alternate TSSs are capable of producing shortened isoforms of proteins (107, 125). Subsequent work has since corroborated these findings. RNA-seq in combination with ribosome profiling has uncovered novel putative translation initiation sites throughout the genome (129), approximately 600 to be exact (130). These sites could have arisen thanks to the ribosome scanning past earlier start codons or from shorter mRNAs, which is where the observed pervasive transcription initiation (107, 125) could come into play. Many of these new sites were located within larger ORFs. Thanks to the compact nature of the viral genome, only a few of these putative translation initiation sites were found within intergenic regions. Some were also found in untranslated leader sequences and on the antisense strand. The predominant start codon was still predicted to be AUG, although other near-cognate start codons were found, particularly for post-replicative genes. The short nature of these additional ORFs did raise questions regarding their biological importance though. Furthermore, questions still remain with regard to the synthesis and stability of the peptides encoded by these ORFs (129). On the other hand, these newly identified ORFs have the potential to play important structural or regulatory roles in VACV replication (130). Regardless, it is evident that POXVs seem capable of encoding more proteins that previously thought. Pervasive transcription and translation initiation serve to increase the coding capacity and expand the functional repertoire by allowing for the generation of extra RNAs and proteins. These findings are not unique to VACV and POXVs. Ribosome profiling has also revealed short ORFs and alternative start codons in studies with cytomegalovirus (CMV), Kaposi's sarcoma-associated herpesvirus (KSHV), bacteriophage lambda, and even uninfected

eukaryotic cells (129, 130). Given that POXVs have acquired cellular genes over the course of their evolution as a result horizontal gene transfer (79, 131, 132), presumably via reverse transcription since introns are missing (130), pervasive transcription and translation initiation may provide an explanation for how these novel ORFs were initially expressed (129, 130), since they would have been randomly inserted into the genome. Then, natural selection should result in enhancement of promoter and translation sites which would allow optimal expression (129). Such genomic plasticity would allow POXVs to adapt to changes in their environments. These newly discovered putative gene products may impact the differences in host range and pathogenicity observed between POXVs, which in turn contribute to the efficacy and specificity of POXVs being used for oncolytic virotherapy and as vaccine vectors (130).

2.6 Poxvirus Evolution

Debates regarding the origins and evolution of viruses have gone on for many years, with many theories being put forth to explain the existence of viruses, which are the most abundant biological entities on the planet and outnumber cells by at least an order of magnitude (133-135). Two of the more common theories are the cell degeneration model of virus origin and the escaped-genes model. Other models do exist though, including: the origin of viruses from a primordial gene pool; an ancient lineage of viruses spanning the three domains of cellular life, based on the presence of a jelly-roll capsid (JRC) protein in various groups of DNA viruses; and at least three ancient RNA virus lineages giving rise to independent DNA viruses lineages, which in turn imparted the ability of DNA replication onto their hosts, RNA cells (43,

133, 135, 136). The ability to accurately and quickly sequence genomes has made it easier to find similarities and trace lineages of organisms and individual genes. A paper by Koonin et al. (133) suggested that the first two models are not supported by comparisons of viral genomes and leaned towards the primordial gene pool model as the most likely origin of the major virus lineages (133, 137). Starting with prokaryotic viruses, this scenario would involve extensive mixing and matching of diverse genetic elements. The next major event proposed by the paper would be the emergence of eukaryotic viruses, in which the newly-emerged eukaryotic cell would serve as a site for extensive mixing and matching of viral and cellular genes (133, 135). The existence of viral genes with no known cellular homologs, such as the superfamily 3 helicase, provides evidence for an ancient pre-cellular virus world (43, 133, 137). This idea has since been explored in more depth and supported by other phylogenetic and comparative-genomic studies (138-144). These findings also do not support another popular theory regarding viruses making up, or originating from a now-extinct, fourth domain of life (140, 145).

Another important point of the 2006 Koonin et al. paper (133) is that it puts forth the idea that viral and cellular evolution are uniquely interconnected (133), outside of the oft-cited "arms race" or Red Queen hypothesis of how viruses and cells drive the evolution of their responses to each other (134, 146). Another level of this interconnectedness is evident when it was realized, not surprisingly, that viruses have acquired, and sometimes repurposed, genes from their respective hosts via a process called lateral or horizontal gene transfer (LGT or HGT) (43, 79, 131-134, 136, 143, 144, 147-154). Perhaps not surprisingly, this mechanism is a two-way street, as

cellular hosts have been known to recruit viral genes for a range of functions (134, 149, 154, 155). In fact, the transfer of genes may be a bit more one-sided than commonly believed, and not in the direction that would be expected, i.e. cell-to-virus. Viruses may have played a major role in the evolution of cells, with cells being the main recipient of new genetic material (154). Adding another layer onto this is the fact that cellular hosts are not the only source of new genes for viruses and can include bacteriophage (79, 135, 144, 152, 153) and other viruses (141, 144, 156), bacteria (136, 144, 152, 153, 157) and other prokaryotes (144), other eukaryotic organisms besides their respective hosts, in the case of eukaryotic viruses (144), diverse families of mobile genetic elements (131, 143, 144, 150, 157-159), and even sources with ambiguous genetic phylogenies (144). For example, some POXVs contain homologs of vertebrate host immune system genes, including multiple members of the interleukin-10 (IL-10) family, which also includes human ILs-19, -20, and -24 (43), (151), while some POXVs contain homologs of the vertebrate vascular endothelial growth factor A (VEGFA) (43). The proteins encoded by the members of the Chordopoxvirinae subfamily generally show greater similarity to eukaryotic proteins than to viral proteins from non-POXVs (160), such as in the case of the viral DNA ligase (152). Evolution can also occur at the species level when a new environment or host is encountered. This can drive and/or be driven by changes at the molecular level. POXVs are no exception to these two phenomena.

2.6.1 Poxvirus Phylogeny

POXV phylogeny is an interesting story, both at the species/strain level and at the gene/protein level. POXVs are part of the NCLDV superclade, first described in

2001 (2, 43). Since then, a new name for the order containing these viruses has been proposed, Megavirales, given their large genomes and viral particles (2). The original members of this superclade were the Poxviridae, Asfaviridae, Iridoviridae, Ascoviridae, and Phycodnaviridae families. Since 2003, the number of virus family members has more than doubled (**Table 2.4**), with the discoveries of the Mimiviridae, Marseilleviridae, Megaviridae, Pandoraviridae, Pithoviridae, and Faustovirus families (2, 4, 161). This extremely broad range of viruses infects hosts across seemingly the entire range of eukaryotic cells. Two of the five currently recognized supergroups of eukaryotes (*Rhizaria* and *Excavata*) have not yet been studied in enough detail to determine if they too are infected by NCLDVs. The perceived ancient origin of the NCLDVs may have been concomitant with eukaryogenesis as well (2).

TABLE 2.1 Current families within the NCLDV superclade (proposed order Megavirales (2))[a]

Family	Yr discovered	Host(s)	Replication site	Assembly site	Genome (kb)
Poxviridae	1798?	Vertebrates, insects	Cytoplasm	Cytoplasm	Linear (130–380)[c]
Asfarviridae	1921	Pigs, warthogs, insects	Cytoplasm	Cytoplasm	Linear (170–190)[c]
Iridoviridae	1966	Fish, frogs, snakes, insects	Nucleus	Cytoplasm	Linear (102–212)[d]
Ascoviridae	1983	Insects, moths	Nucleus	Cytoplasm	Circular (157–186)
Phycodnaviridae	1981	Algae	Nucleus	Cytoplasm	Linear (100–560)
Mimiviridae	**2003**	**Amoebae, zooplankton**	**Cytoplasm**	**Cytoplasm**	**Linear (~1,200)**
Marseilleviridae	**2009**	**Amoebae**	**Cytoplasm**	**Cytoplasm**	**Circular (368)**
Megaviridae	**2010**	**Amoebae**	**Cytoplasm**	**Cytoplasm**	**Linear (1,208–1,259)**
Pandoraviridae	**2013**	**Amoebae**	**Cytoplasm**	**Cytoplasm**	**Linear (1,900–2,500)**
Pithoviridae	**2014**	**Amoebae**	**Cytoplasm**	**Cytoplasm**	**Linear (610)[d]**
Faustovirus	**2015**	***Vermamoeba vermiformis*[b]**	**Cytoplasm**	**Cytoplasm**	**Circular (455–470)[e]**

[a]Those families considered to be giant viruses, discovered starting in 2003, are shown in bold. Classification and tree topology are still developing, with, for example, the recently discovered *Dinodinavirus*, *Faustovirus*, *Cedratvirus*, *Kaumoembavirus*, and *Mollivirus* also being considered members of the NCLDV superclade.
[b]A protist.
[c]Has covalently cross-linked ends and inverted terminal repeats.
[d]Circularly permuted and terminally redundant. The upper size limit is 303 kb if redundancy is included.
[e]Eight out of nine *Faustovirus* genomes were circular (129, 130). Adapted Table 1 from Mirzakhanyan, Y. and P. D. Gershon. Multisubunit DNA-dependent RNA polymerases from vaccinia virus and other nucleocytoplasmic large-DNA viruses: impressions from the age of structure. Microbiol Mol Biol Rev, 2017. 81: e00010-17. (4). Reprinted with permission under the Creative Commons Attribution License.

All NCLDVs have: representatives of a set of 47 genes, more or less (2, 143, 144, 162), needed for key viral functions, presumably conserved from the ancestral virus; a common virion architecture; and common major biological features, specifically replication within cytoplasmic factories. All the NCLDVs share five core genes, encoding: the major capsid protein (POXV D13 gene), a helicase-primase (D5), the DNA polymerase elongation subunit family B (E9), a DNA-packaging ATPase (A32), and the viral late transcription factor (A2). All the NCLDVs form large, icosahedral capsids which consist of a single, homologous double B barrel jelly roll protein. The only exceptions are POXVs, which form a unique brick-shaped virion, and Ascoviruses, which have allantoid capsids. It should be noted, however, that POXV IVs resemble the icosahedral shape of the other members of this group. All NCLDVs replicate either entirely in the cytoplasm of infected cells or start in the nucleus and end up in the cytoplasm (2). The viruses that do the latter are the Iridoviridae, Ascoviridae, and Phycodnaviridae families. All the others found and classified so far are cytoplasm-only (4), although, in the case of the Mimivirus at least, it is not clear if the nucleus is involved in genome replication (163).

Members of this superclade also temporally regulate their gene expression, with distinct phases separating each one for those that have multiple. These phases are generally broken down into three classes: (immediate/delayed) early genes, intermediate genes, and late genes. Such regulation is possible using distinct promoter regions for each class of genes, which differ from one another but exhibit a pattern of conservation within the same group. This control over gene expression provides a clear evolutionary advantage and indicates that these sequences were

selected to ensure each gene is transcribed when, where, and at the level required. Also of note is that the promoter regions tend to be AT-rich, even in genomes with a high GC content, suggesting that the ancestral virus likely had an AT-promoter sequence (161). Such conservation of the promoter sequences seems to support previous studies which found that, in viruses, sites outside coding regions tend to be more constrained than synonymous sites within the coding regions, due to the role the former has in regulating gene expression (43). Most members of this superclade encode their own MSDdRPs, which tend to resemble their cellular counterparts. The viral versions, however, generally appear to be stripped of their higher-order cellular control and honed for speed and processivity (4). Different studies seem to point to the fact that POXVs are more closely related to Asfavirus (ASFV) than any of the other NCLDVs, as they use similar transcriptional strategies (161) and have their own unique virion structures (2, 138). As a result, when phylogenetic trees of the NCLDVs are constructed, even though each family forms its own clade, the tree has two obvious branches: with POXVs and ASFV on one and the rest of the NCLDVs on the other (2, 138, 164), suggesting a more recent common ancestor of the two families (161, 164).

2.6.2 History of VARV

Many groups have tried to determine the origin and history of VARV for academic purposes but also to understand how and why viruses move and change over time, perhaps enabling them to come away with lessons that can be used for present-day viruses, especially other POXVs (1, 8, 9, 15). Fortunately for POXV researchers, smallpox was such a plague among humanity that there are numerous

historical medical records documenting cases, providing some evidence to help chart the movement of the disease as it spread across the globe (1, 8, 9, 15). The earliest and most reliable description of the disease dates back to the 4th century AD in China (1, 9, 15, 43) , then India in the 7th century, and in Southwest Asia and the Mediterranean in the 10th century (9). The advent of high-throughput sequencing and the generation of complete sequences for entire POXV genomes has aided a great deal in the phylogenetic analysis of VARV lineage (9).

It is believed that the etiological agent of smallpox occurred no earlier than 10,000 years ago (46), with VARV having a relatively recent origin of about 3300 years ago (1, 8, 44-46), This timeline coincides with the split of VARV, CMLV, and TATV from a common ancestor (43-45), presumably a CPXV-like virus (1) that infected rodents (8, 46) but had a broad host range (1, 44). All of these viruses are closely related, only infect one particular host (1, 46), and VARV and CMLV have high mortality rates (1, 44). Such an occurrence could only have taken place when this ancestral virus encountered new, susceptible hosts in the same area and at a sufficient density and size to support the spread of the virus(es) (8), (44), (1), (45). Once this happened, the disease could then persist in the population by infecting people in surrounding areas and then return to a previous area once a large enough new crop of susceptible individuals has been introduced (1, 8). One event in particular was keyed in on by Babkin and Babkina (1): the Santorini Eruption, which was one of the largest volcanic events in recorded history and caused considerable changes in the climate (1). This event would have driven camels and naked sole gerbils, the only known hosts for CMLV and TATV, respectively, together in the

Horn of Africa no earlier than 4000 years ago, consistent with phylogenetic analysis regarding the emergence of these viruses (1, 44). The specific event believed to have triggered this divergence was the introduction of the camel, which represented a new potential host with a unique antibody structure and a change to which the ancestral virus had to adapt (1, 44).

Along with climate change (1, 44), the most widely-accepted drivers of pathogen emergence are destruction of the environment of potential hosts, introduction of the pathogen into a new area, spread of the pathogen to other populations of hosts, and interactions of the pathogen with the host immune system. The authors were cautious though and acknowledged that just because this cataclysmic event correlates with the emergence of these viruses does not mean it caused it (1).

Another factor that probably coincided with and may have helped this speciation was the disruption of the ATI gene in VARV. Deletion of this gene in CPXV enhances OPXV respiratory infection, which is the primary method of spread for VARV in humans and MPXV in rodents, which also has a disrupted ATI gene. Such an occurrence would seem to provide a selective advantage for these viruses. In fact, retention of a functional copy of this gene may have negatively affected the abilities of these viruses to spread, as ATIs would seem to be more beneficial if the host population is sparse or spread is inefficient or indirect (94). A breakthrough in charting the molecular evolution of VARV was achieved when restriction fragment length polymorphism (RFLP) analysis revealed that the West African and South American strains formed a separate subtype, with drastic differences in their genomic

organization when compared to other VARV geographical variants. Even more intriguing is the fact that these two strains within this subtype form two distinct phylogenetic groups, suggesting that they had evolved independent of each other (1, 46). The divergence of the West African strain happened about 570 years ago (1, 46). It then evolved into the South American strain about 350 years after it was imported to the continent (8, 46) as a result of the slave trade (46). Over the course of this evolution, the case fatality rate (CFR) dropped from 8-12% to less than 1% (1, 44).

Parallels have been drawn between VARV, MYXV (1, 44, 45), and ECTV (45), which infect European rabbits and mice, respectively as their non-natural hosts. The mortality rates can approach 100% in both of these cases (45). Such high lethality rates are indicative of viruses which have only recently adapted to new hosts (1, 44), prior to genetic changes attenuating the virus (45), and this property provides further evidence of the fairly recent emergence of VARV (1, 44). After introduction into a new host, viruses usually undergo evolutionary changes which reduce the CFR, as has been witnessed with MYXV in Australia (1, 45), and with MPXV, between the West African and Central African strains, where the former strain lost the complement binding protein, resulting in a reduction in the CFR compared to the latter strain (1, 44). The region of West Africa provides a fascinating example of how geography can influence the emergence of a pathogen. The emergence of West African VARV and MPXV mostly likely occurred due to the geographical isolation of the region from the rest of the continent due to the presence of the rain forest. This region is also separated from Central Africa and South Africa by the Adamawa Mountains to the east and by the Sahara Desert to the north, respectively. The

distinct separation of virus families into geographical subtypes is not limited to POXVs. Examples of viruses specific to West Africa have been found in subtypes of Ebola virus, human immunodeficiency virus (HIV) -2, simian immunodeficiency virus (SIV), and hepatitis C virus (HCV) genotype 2 (46). Parallels have also been made between VARV and Measles, which is also specific to humans in terms of pathogenicity (1). Measles, which is theorized to have evolved from a rinderpest virus 1000-1500 years ago (45), requires a population of about 200-300 thousand people in order for a sufficient number of susceptible children to be introduced regularly so that it can be maintained in the population (1).

2.6.3 History of VACV

The history of VACV is more convoluted, as has been mentioned before. This has been due to multiple passages of viruses in different animals and the use of both cows and horses as the source material for protection against smallpox (13, 16, 24). In fact, prior to the start of the smallpox eradication campaign in 1967, smallpox vaccines from multiple sources were used. Once the global effort began, however, the WHO standardized four VACV strains for use as vaccines (14, 24). For nearly a century and a half, however, it was assumed that CPXV was the virus used as the vaccine against smallpox, but it was not until 1939 that a British researcher named Allan Downie demonstrated that VACV and CPXV, though related, were not the same virus (14), (13, 16, 24, 43). Furthermore, VACV is neither an attenuated strain of VARV nor a recombinant of CPXV and VARV (24). It has also been shown that all modern (20^{th} century) vaccine strains are composed entirely of VACV (16, 43).

A HSPV-like virus is now assumed to be the origin of VACV (1), though this is unclear. Recent evidence however supports the idea that at least early smallpox vaccines were based on a HSPV-like virus, as a recently-found 1902 vaccine was closely related to HSPV (13, 14). HSPV, which was not present in the Americas and was only found in Europe (13, 14, 24), is believed to have diverged from its ancestor only about 3000 years ago (44). VACV and HSPV are closely related and HSPV is now believed to possibly be the ancestor or at least related to the ancestor of the VACV lineage (13, 14, 16, 24), including the Brazilian vaccine strain known as VACV IOC and the two feral or field strains, CTGV and Serro 2 (12, 16). Horsepox as a disease is rare and is now believed to be extinct in nature, at least in Europe (13, 16, 24), with the only strain currently known being isolated from Mongolian horses in 1976 (13, 16, 23, 24). Interestingly, the genome contains genes that are fragmented in some VACV strains but has gene fragments of genes that are intact in other VACV strains, suggesting that multiple HSPV strains may have been the ancestors of different VACV lineages (13, 23). A recent paper, however, detailed the generation of live synthetic chimeric horsepox virus (165), raising the possibility of studying the virus and others like it while also raising serious ethical and biosecurity concerns to some in the field (166-173). This history of unintentional evolution of the vaccine strains has created pools of highly diverse quasispecies. On a phylogenetic tree though, three obvious clusters form: the American or Dryvax cluster, the South American cluster, and the Eurasian cluster , all of which are separate from both CPXV and VARV (12, 16). HSPV clusters with the Brazilian strains of VACV

using multiple clustering methods, indicating how closely related these strains are and lending credence to the idea that HSPV may be the ancestor of VACV (12, 16).

2.6.4 Accordion Model, Poxvirus Adaptation, and Experimental Evolution

In the field of virology, it was generally believed that the polymerases of RNA viruses are more error-prone than those of DNA viruses. The two sets of enzymes actually have similar fidelities. What sets them apart is that the polymerases of DNA viruses have proofreading functions, so fewer mutations accumulate (174, 175). Additionally, some DNA viruses, like Herpesviruses and POXVs encode proteins involved in DNA repair (175). This means that DNA viruses mutate and thus tend to evolve on a slower scale than their RNA counterparts (46, 160), with the ssRNA viruses having a mutation rate of $10^{-2} – 10^{-4}$ substitutions/site/year (46) and HIV-1 having a spontaneous mutation rate of about 2×10^{-5} mutations/bp/replication. In general, RNA viruses tend to have higher rates of mutations (43) and thus higher genetic variability (160). What is becoming clearer though is that DNA viruses possess other ways of adapting to changing environments than simply relying on the mistakes of their polymerases. Even among DNA viruses though, POXVs have not gotten much attention in this regard until recently. For instance, more effort had been devoted to estimating the substitution rates of herpesviruses (43). That said, it has now been calculated that OPXVs have a mutation rate of about 10^{-8} substitutions per nt per replication cycle (175). This is in line with the calculated mutation rates of DNA viruses, on the order of 10^{-7} to 10^{-9} mutations per site per round of replication, as they tend to have substitution rates closer to those of their hosts (160).

Viruses make great models for studying evolutionary processes and evolution at the population level, due to their short generation times and high mutation rates, especially in the case of RNA viruses. Genome sequencing advances and reductions in the cost of the process, in combination with short viral genomes, have made it easier to examine viral responses to both natural and artificial selection pressures. POXVs, due to their being DNA viruses and having fairly large genomes, have not been extensively used for population studies (43), although this has been changing (69, 122, 146, 174, 175).

2.6.5 Genome Accordion Model

POXVs, along with the other members of the Megavirales order, or the Giant Viruses (GVs) as they are also called (144) , are believed to have obtained their large genomes through successive steps of genome expansions, in the form of gene duplications and gene transfers, balanced by genome reductions, in the form of gene deletions (42, 56, 58, 143, 144, 151), with additional supplementation through the movement and amplification of diverse genetic mobile elements (143). This multistep process is referred to as the accordion model of genome evolution. In general, there does not seem to be a trend toward either genome expansion or genome contraction, as each expansion is balanced out by a reduction, at least for certain families of GVs (144). Many pieces of evidence support this model of genome evolution for these viruses (69, 143, 144, 146, 174). Additionally, it is possible that this process is not unique to the NCLDVs and may extend to all large DNA viruses and phages (143). These findings also support the theory that genome gigantism occurred early in the evolution of this virus superclade, before the divergences of the

different GV lineages (144). It should be noted again, however, that all the current members of the OPXV genus appear to be an exception to this rule regarding the overall balance of genome size and/or content. None have acquired or generated a gene not found in CPXV, with the lone exception being the short interspersed element (SINE) found in TATV (176), supporting the idea that CPXV most closely resembles the last common ancestor of this genus. This finding also suggests that the members of this genus have undergone genome reductions, via amino acid changes and gene fragmentations and deletions, since that divergence, as they adapted to new hosts and environments (7, 42, 56, 176, 177).

2.6.6 Nucleotide Composition and Codon Usage

POXVs also have complex evolutionary processes acting upon their nucleotide composition and codon usage. The nucleotide composition of POXVs seems to influence their codon preferences. Interestingly enough, the codon usage of POXVs seems to be genus-specific, adding yet another layer to how evolution can act upon these viruses. Specifically, the OPXV genus virus members analyzed in the 2011 study by RoyChoudhury et al. (177) all have AT-rich genomes. Viruses with AT-rich genomes show an overall preference for A and T in the first two codon positions but a much stronger preference for A and T at the third codon position. The POXVs with GC-rich genomes on the other hand preferred GC at the first codon position, AT at the second codon position, but strongly preferred GC at the third codon position. This supported the finding that there was a strong and significant correlation between overall nucleotide composition and the nucleotide composition of the third synonymous codon position. Such a pattern implied that viruses

phylogenetically close to each other, particularly those within the same genus, share similar nucleotide compositions. These findings are interesting since the three positions are subjected to different evolutionary forces. The first two positions are determined by selective constraints, while the third position is supposedly primarily determined by mutational pressures. As such, it appears the third position may have greatly influenced the codon usage in POXVs, regardless of the overall nucleotide composition of the genome in question. The authors also felt that this begged the question of whether mutational bias might have superseded the forces of translational selection during the evolution of codon bias in the Poxviridae family. The third codon position preference led the authors to conclude that GC-rich POXVs probably evolved under much greater mutational pressures than AT-rich POXVs (177).

All these evolutionary forces combined to largely determine the existing set of preferred codons for each GC- and AT-rich genus and may have ultimately led to a path for optimal fitness for the POXV genetic machinery. Other factors such as gene length and expression level also appeared to affect codon usage patterns to some degree. Sequence length directly affected the codon usage bias of all GC-rich genera and a few AT-rich genera, a group which does not include the OPXV genus. In general, though, GC-rich POXVs have longer genes with higher codon bias, while the opposite is true for AT-rich POXVs. The expression levels of genes showed a similar trend, in that genes with higher levels of expression in the GC-rich genera exhibited more codon usage bias while genes with lower levels of expression in the AT-rich genera showed more codon usage bias. Not surprisingly these trends carryover to dinucleotide usage as well. AT-rich genera are overrepresented by the dinucleotides

TA/AT and AA/TT and underrepresented in GC-rich viruses. The opposite is true for GC-rich viruses, with GG being an exception to the rule. Again, not surprisingly, AT-rich genera viruses preferred codons only which ended in A or U, while viruses from GC-rich genera only used codons with either a G or C at the third synonymous position (177).

2.6.7 Gene Content

Nucleotide changes have also been examined with regard to their effects on genome evolution and gene content, especially in the context of the OPXV genus. Specifically, the reductive evolution observed in this genus was found to be the result of single nucleotide polymorphisms (SNPs) and insertions/deletions (indels). The accumulation of such mutations led to individual missense mutations and gene truncation and fragmentation via the introduction of early stop mutations (ESMs). These ESMs would obviously result in the loss of function of the protein in question and possibly the eventual elimination of that particular ORF. Perhaps counterintuitively, this study by Hatcher et al. (176) found that out of all isolates tested, those with the shortest genomes contained the greatest number of deleted genes, deteriorating genes, and highest concentration of ESMs. This may be because of a change in the selective pressures acting on the virus, such as when it encounters a new host. Proteins that are no longer required for replication in a new host will be subjected to the accumulation of numerous mutations. In fact, if a once-required protein becomes deleterious to the virus in its new setting, it will be selected against and gradually degraded until it is removed from the genome. This will also save the virus from the burden of replicating nonfunctional DNA. Such a mechanism has been

proposed to be an important step in narrowing the host range of a virus, such as in the cases of VARV, ECTV, and MOCV. Similar phenomena have been witnessed in symbiotic and obligate parasitic bacteria, which have smaller genomes than their free-living counterparts. From a research standpoint, such occurrences aid in determining when selection pressures are lost, as this mutation rate approaches the error rate of the viral DNA polymerase in the absence of selection pressure. Despite the truncation and fragmentation of certain genes, it has been shown that many of them are still transcribed because promoters are often untouched or only minimally changed in genes that are in the beginning stages of degradation. The authors point out, however, that this does not mean these transcripts are necessarily translated. Even if they were, these proteins may not be functional. The loss of its full repertoire of genes may provide a virus with the potential to maximize its replication in one or a few hosts, but it also runs the risk of encountering an evolutionary dead end. In contrast, a virus such as CPXV that maintains a full complement of genes can maintain its ability to infect a more diverse range of hosts. That being said, the authors warn that recombination between closely related OPXVs could occur and thus provide the genetic diversity needed to escape a potential evolutionary dead end, leading to the creation of novel host-specific pathogens (176).

2.6.8 Adaptive Molecular Evolution

Any genes and proteins still needed by the virus will more than likely continually be under selective pressures, with some genes/proteins experiencing greater selection than others. As mentioned before, POXVs have a linear genome with genes essential for virus replication in the center. This central region is highly

conserved across the entire POXV family (7, 178), both in terms of gene content and sequence (178). The arrangement of these core genes is similar in ChPVs (7) and in OPXVs these genes form a continuous block in the middle of the genome (178). Genes on the terminal ends, however, are much more variable and diverse, as they tend to be involved in host interactions (7), (178). Many of these are host species specific, meaning they have adapted to that particular host's immune response. Such genes include immune evasion genes that block apoptosis, inhibit cytokines, or inhibit the interferon (IFN) response. A lot of these genes tend to be non-essential for the virus to replicate in cell/tissue culture, but virulence factors required for mediating interactions within its natural host. Because of this, many of these genes are highly specialized and thus differ between different genera, between species within a genus, and even between different strains within the same species (178). There are three types of selection that can occur regarding the ratio of non-synonymous to synonymous mutation fixation rates: purifying selection, neutral selection, and diversifying selection. In purifying selection, there is a higher rate of synonymous mutations compared to non-synonymous mutations, implying that there are stringent functional and/or structural constraints on a particular amino acid. The highly-conserved nature of the core region of the POXV genome would suggest that this region is under purifying selection, as non-synonymous mutations are more likely to have deleterious effects on the virus. In neutral selection, the two mutation rates are equal and a change to a different amino acid at a certain residue has neither a positive nor a negative effect on the protein. In diversifying selection, otherwise known as adaptive molecular evolution, non-synonymous mutations become fixed in the

population at a greater rate than synonymous mutations. Genes at the termini of the POXV genome, which would be involved in host-pathogen interactions are more likely to be under this kind of selection as they would be more amenable to change and thus possess greater sequence diversity than the central region (178).

A paper by Esteban and Hutchinson (178) set out to investigate the selective pressures that act on the different regions of the POXV genome and see what role adaptive molecular evolution has played in the evolutionary history of POXV genes. They reasoned that the process of adaptive molecular evolution is a key mechanism for species divergence to take place and that finding proteins and/or specific residues under this kind of selection may aid in understanding gene function, identifying essential functional regions in genes of either known or unknown function, and categorizing genes of unknown function as ones involved in host interactions. Using the CPXV-BR (Brighton Red) strain as a model, what they found was that, out of the 175 gene families examined, 79 were found to be under diversifying selection, with 20 of these having an as-of-yet unknown function. Many of these 79 genes were, not surprisingly, found on the terminal ends of the genome and were modifiers of the host response to infection. What was a bit surprising though was that genes involved in viral replication and virion structure were also demonstrated to be under diversifying selection. The authors theorized that this may be due to these protein products being packaged into the virus particle. Some of these proteins included A4, a major core protein and the most abundant protein in the virion (by weight), and A7, D8, and B5. This finding implicated a greater role for diversifying selection in host-interacting genes, as it means this mechanism is not just limited to immunomodulatory genes but

encompasses genes whose products interact with the host, such as major antigens (178).

Furthermore, their analysis found that multiple POXV genes in the central region of the genome were also experiencing diversifying selection, which was not expected, as these genes are not typically considered when determining virulence. Some of these same gene products are packaged in the virion and thus may encounter the host antibody response, suggesting that non-immunomodulatory proteins may also contribute to virulence and that interactions with the host serve as the basis for adaptive molecular evolution. Therefore, going forward, they felt it may be prudent to take adaptive molecular evolution into account in well-conserved genes in addition to major genomic differences when attempting to explain differences in virulence between strains of POXVs. Such considerations may also prove useful when examining the potential for POXVs to adapt to new hosts, such as MPXV in humans. They concluded that diversifying selection plays an overall important role in POXV evolution and plays a more important role in the ends of the genome versus the center, though the latter should not be discounted. Additionally, they felt their findings of residues under diversifying selection would provide groundwork for future mutagenesis studies related to determining if these sites are important for host-specific interactions, like protein-protein contacts or immune epitopes. This would be particularly useful when dealing with genes that have unknown or poorly-defined functions, as these data would provide a jumping-off point for characterizing these proteins as potential host interaction molecules or virulence factors and possibly

highlighting highly-conserved functional or diverse host-specific residues that would be of special interest (178).

2.6.9 Experimental Evolution

Up until now, all the methods discussed regarding POXV evolution have been primarily in the context of natural selective pressures. These mechanisms all function in laboratory settings, under artificial selection, in a technique known as experimental evolution. This allows researchers to study evolution in a controlled environment, so they can monitor when viral fitness begins to improve in a given scenario and under what conditions. The advent of high-throughput sequencing, coupled with its decreasing cost, makes it easier to follow said changes in near real-time. This has made it much more feasible to test the genome accordion model of POXV evolution (143, 144) under carefully regulated conditions (69, 146, 174, 175). The first study, by Elde et al. (146), utilized a mutant VACV with the E3L gene deleted. E3 is an antagonist of protein kinase R (PKR), one of the most potent innate defense sensors, which becomes activated upon detection of dsRNA in the cytoplasm. This placed tremendous selection pressure on the VACV K3L gene, another antagonist of PKR. What they saw was that over the course of the experiment, multiple copies of the K3L locus were made and the genome expanded but that by the end only one copy remained. This retained copy, however, had a mutation in it, H47R, which allowed K3 to better combat PKR and thus increase viral fitness (146). Adaptive mutations don't have to necessarily arise in the amplified locus, however, as Brennan et al. (69) showed in their paper. Their study also looked at PKR and its antagonists. The difference was they used a weak antagonist called rhtrs1, a gene encoded by rhesus

cytomegalovirus, in a VACV which had neither E3L or K3L. Their findings revealed adaptive mutations in the A24R and A35R loci, neither of which had previously been implicated in PKR antagonism. A24 is the catalytic subunit of the vRNAP, while less is known about A35. What is known about A35 is that it is highly conserved in the OPXV genus and reportedly involved in modulation of the adaptive immune response. These adaptive mutations arose at the same time the then-amplified rhtrs1locus was collapsing back into a single copy (69). Similar results were obtained by Cone et al. (174). In their experiments, they also used a VACV E3L deletion mutant, thus exerting great selection pressure on the K3L gene, just like Elde et al. (146). This also resulted in a genome expansion involving the K3L locus (174). Unlike the results of the Elde et al. study, however, Cone et al. found adaptive mutations outside of K3, similar to Brennan et al. (69). Unlike the Brennan et al. study though, the adaptive mutations found by Cone et al. were only located in the A24R locus.

Such localized genome expansions allow viruses that have lower mutation rates, such as POXVs, to sample mutational space while allowing for temporary enhancement of gene expression (69, 146, 174). Higher levels of expression are particularly useful in the case of a host-range gene such as K3L (69, 146, 174), and would confer a fitness advantage early in infection in which selection pressures are introduced, such as when a virus encounters a new host. Once beneficial mutations arise, the subsequent contraction reduces the burden on the virus to replicate a larger genome. In fact, during later stages of infection, viruses with expanded genomes may be at a disadvantage, especially when up against viruses carrying the adaptive

mutation (69, 146, 179). Thus, this fairly straightforward mechanism of recombination-driven genomic expansions and contractions allows for more rapid viral evolution and adaptation than if said viral populations had to rely solely on their relatively low mutation rates (146). Experimental evolution enables researchers to catch these transient genome expansions, which can be selected against in a population once it is no longer beneficial, and thus never discovered (146, 174). This can be crucial since gene amplification and collapse is reportedly a dynamic process, at least in bacteria (69). These studies highlight the fact that gene amplification or gene copy number variation (CNV), otherwise known as structural variation (174, 175), plays an important role in adaptation of viruses with low mutation rates (69, 146, 174). This strategy of using CNV coupled with nucleotide substitution may also be utilized by other viruses with low mutation rates and somewhat limited restrictions on genome size, such as other large DNA viruses and phage (143, 146, 179). Gene amplification followed by subfunctionalization or neofunctionalization of the gene copies is highlighted by the existence of families of paralogous genes in numerous DNA virus families (146, 179), via a process known as lineage-specific gene expansion (143). Transient CNV can accelerate the fixation of mutations that provide a modest fitness advantage or even a potential fitness trade-off. Thus, CNV can enable POXV adaptation, and virus evolution as a whole, through both direct and/or indirect means (174, 179). Gene amplification is not just limited to viruses, however, and has been reported as a mechanism common to prokaryotes, archaea, and eukaryotes (69, 179). Researchers can take advantage of this strategy, especially when employed by viruses, to identify functionally important amino acid residues, in

the case of adaptations within the amplified locus (146), and to identify potential novel host-range determinants and virulence factors, in the case of adaptations occurring outside of the amplified locus (69, 174).

CNV is not the only mechanism by which POXV can adapt however. As beneficial as it might be, it has yet to be identified in strains isolated out in the field, probably due to its transient nature (146, 175), and so far has only been confirmed in laboratory settings using cell culture and high selection pressures. In fact, studies looking at POXV evolution have focused exclusively on changes at the genomic level. One study set to examine POXV adaptation at the level of the proteome using CPXV as their model. This study, by Grossegesse et al. (175), found that, overall, the genome was relatively stable during passaging. While they found minor genomic variants, only one increased in frequency in both cell lines used. This change was found in the A25R gene (A24R in VACV) and its increase in frequency correlated with the observed increase in viral fitness. What they did find was that most of these variants mapped to the tandem repeats, or microsatellites. These are found throughout the POXV genome and can account for about ¼ of the sequence. ESMs have been found to accumulate in ChPV microsatellites, leading to the theory that microsatellite hypervariability is a major source of POXV genome variability, and thus, a source of POXV adaptation. The authors felt their data supported the idea that microsatellite hypervariability is a mechanism by which CPXV adapts, at least in cell culture. Their primary finding, however, was increased amounts of certain viral proteins over the course of the experiment. They did admit though that this could be due to increased expression, greater abundance in virions, or some combination of

both. The proteins identified could be functionally categorized in one of three ways: membrane-associated, immune evasion, and diverse enzymatic activities. All the genomic variants were in genes involved in transcription, leading the authors to conclude that the proteomic changes were regulated at the level of transcription. Their findings implicate proteomic changes as a novel mechanism for POXV adaptation, which may also apply to other NCLDVs. The authors did caution however that their study was only descriptive in nature and that the association between the observed proteomic changes in the presence of viral fitness changes was hypothesized due to the correlation between the two phenomena.

Chapter 3: RNA polymerase mutations selected during experimental evolution enhance replication of a hybrid vaccinia virus with an intermediate transcription factor subunit replaced by the myxoma virus ortholog[1]

3.1 Summary

High-throughput DNA sequencing enables the study of experimental evolution in near real time. Until now, mutants with deletions of nonessential host range genes were used in experimental evolution of vaccinia virus (VACV). Here, we guided the selection of adaptive mutations that enhanced the fitness of a hybrid virus in which an essential gene had been replaced with an ortholog from another poxvirus genus. Poxviruses encode a complete system for transcription, including RNA polymerase and stage-specific transcription factors. The abilities of orthologous intermediate transcription factors from other poxviruses to substitute for those of VACV, as determined by transfection assays, corresponded with the degree of amino acid identity. VACV in which the A8 or A23 intermediate transcription factor subunit gene was replaced by the myxoma (MYX) virus ortholog exhibited decreased replication. During three parallel serial passages of the hybrid virus with the MYXA8 gene, plaque sizes and virus yields increased. DNA sequencing of virus populations at passage 10 revealed high frequencies of five different single nucleotide mutations in the two largest RNA polymerase subunits, RPO147 and RPO132, and

[1] Adapted from: Stuart, C.A. et al. 2018. RNA polymerase mutations selected during experimental evolution enhance replication of a hybrid vaccinia virus with an intermediate transcription factor subunit replaced by the myxoma virus ortholog. J Virol 92(20): e01089-18. doi: 10.1128/JVI.01089-18

two different Kozak consensus sequence mutations predicted to increase translation of the MYXA8 mRNA. Surprisingly, there were no mutations within either intermediate transcription factor subunit. Based on homology with *Saccharomyces cerevisiae* RNA polymerase, the VACV mutations were predicted to be buried within the internal structure of the enzyme. By directly introducing single nucleotide substitutions into the genome of the original hybrid virus, we demonstrated that both RNA polymerase and translation-enhancing mutations increased virus replication independently.

3.2 Introduction

The Poxviridae comprise a large family of viruses that infect vertebrates and invertebrates (40). During their evolution, chordopoxviruses segregated into 11 recognized genera as well as additional unassigned species. Analysis of the genomes of representatives of the various genera showed that approximately 90 genes encoding proteins for essential functions, including entry, transcription, genome replication, disulfide bond formation, and virion assembly, have been preserved (57). A similar number of less well conserved genes are unnecessary for replication in cell culture; many of these genes are involved in host interactions and are present in only a subset of poxvirus genera (180). The diversity of the latter genes is likely related to their acquisition and adaptive modification during the long period of poxvirus evolution and speciation in various hosts (181). In some cases, these viral proteins counteract cellular innate immune responses, whereas others have roles that are not yet understood (182).

In contrast to the relatively low rate of natural evolution within a single host species, propagation of vaccinia virus (VACV) as the smallpox vaccine in unnatural hosts such as calf skin and cell culture over the past 200 years has promoted rapid changes (183). A striking example of this is modified VACV Ankara (MVA), which lost approximately 15% of the genome and suffered a severe host restriction during >500 passages in chicken embryo fibroblasts (184). Presently, high-throughput sequencing methods enable investigation of the experimental evolution of poxviruses in near real time. Thus far, such studies have been limited to VACV host range mutants with deletions of genes involved in evasion of the PKR/eukaryotic initiation factor 2α (eIF2α) antiviral pathway (69, 146, 174, 185). The genetic changes comprised both copy number amplification and point mutation of individual viral genes. We are seeking to extend experimental evolution to essential genes encoding proteins that have coadapted with other viral proteins during natural selection. Our approach is to swap orthologous genes from distantly related poxviruses, thereby encouraging the selection of adaptive mutations that enhance replication. In principle, the results could provide insights into protein interactions and pathways. In this first effort to apply orthologous gene swapping for experimental evolution of poxviruses, we targeted the transcription system.

Poxviruses have three stages of gene expression: early, intermediate, and late (186). An eight-subunit DNA-dependent RNA polymerase (Pol) that is homologous to the polymerases of archaea and eukaryotes (132, 187) acts in conjunction with stage-specific transcription factors and promoters. The VACV early transcription factor is a heterodimer of an 82-kDa and a 70-kDa subunit with ATPase activity that

binds to the core region of early promoters and DNA downstream of the RNA start site, thereby altering the DNA conformation (188-190), and to the RNA Pol-associated protein RAP94 (110, 191). The intermediate factor is a heterodimer of 34- and 45-kDa proteins (117), and the late factor is comprised of 17-, 26-, and 30-kDa proteins (115). The intermediate and late transcription factors have no known catalytic activities, nor is there evidence of direct association with DNA or RNA Pol.

The likelihood that chordopoxviruses use a common mechanism for mRNA synthesis can be inferred from the conservation of their transcription apparatus and from the sequence similarity and functional interchangeability of their promoters (70, 192-194). However, the compatibility of orthologous transcription factors has yet to be investigated. Transient expression experiments described here indicated that orthologous intermediate transcription factors exhibit a range of compatibilities with VACV. Using this information, we constructed hybrid viruses in which the genes encoding the VACV 34-kDa (A8) or 45-kDa (A23) intermediate transcription factor subunit were individually replaced with the myxoma virus (MYXV) ortholog (abbreviated MYXA8 and MYXA23, respectively) and then carried out serial passaging of the poorly replicating viruses to allow experimental evolution and adaptive selection. Substantial increases in replicative abilities of the MYXA8 VACV were associated with mutations in RNA Pol subunits and sequence changes around the translation initiation site of the MYXA8 gene.

3.3 Results

3.3.1 Transient expression assays

To initiate this study, we compared the A8 and A23 intermediate transcription factor orthologs of Chordopoxvirus genera. For diversity, we selected the genes from Myxoma virus (MYXV), Canarypox virus (CNPV), Nile Crocodilepox virus (CRV), and Salmon Gill poxvirus (SGPV), which encode proteins that vary in amino acid identity from 65% to 23% relative to VACV sequences (**Table 3.1**). The order of similarity to VACV was MYXV > CNPV > CRV > SGPV, consistent with their overall genome sequence similarities.

Table 3.1 Comparison of intermediate transcription factor orthologs

Virus	A8 % amino acid identity	A23 % amino acid identity
Vaccinia	100	100
Myxoma	65	58
Canarypox	38	51
Crocodilepox	31	40
Salmon gill pox	29	23

Our plan was to analyze intermediate gene expression after infecting cells with a recombinant VACV lacking the gene encoding A8, A23, or both (ΔA8, ΔA23, and ΔA8/ΔA23 strains, respectively) and transfecting DNA encoding corresponding orthologs. Because the A8 and A23 genes are essential, deletion viruses were previously constructed with the aid of a complementing cell line that expressed both intermediate transcription factors (118). We made new versions of these viruses that constitutively express the bacteriophage T7 RNA polymerase so that a T7 promoter could be used for expression of A8 and A23 genes from transfected plasmids. The starting virus vTF7-3 (126) contains the A8 and A23 genes from the VACV Western Reserve (WR) strain and will be referred to here as the wild type (WT). The VACV A8 or A23 gene or both genes were deleted and replaced by the green fluorescent protein (GFP) or Discosoma sp. Red fluorescent protein (DsRed) open reading frame (ORF) regulated by a VACV promoter to facilitate isolation in the complementing cell line (**Fig. 3.1A**). Loss of the A8 and A23 ORFs was confirmed by PCR, and the inability of the viruses to express viral intermediate proteins was demonstrated by Western blotting (data not shown).

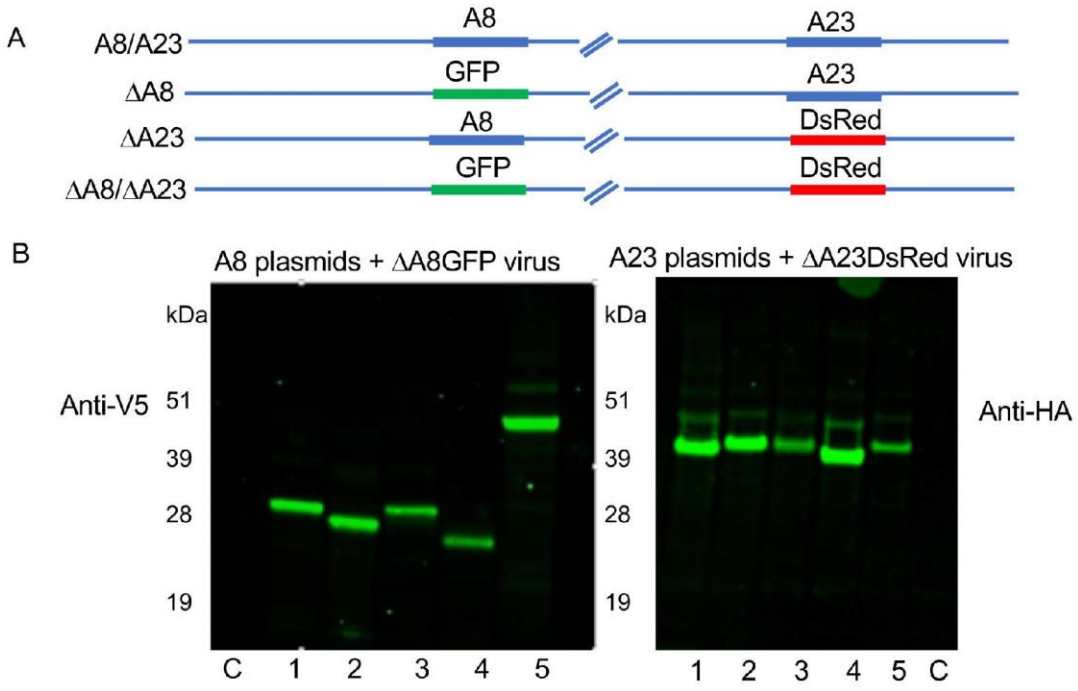

Figure 3.1 Transient expression of intermediate transcription factor orthologs. (**A**) Recombinant VACV that express T7 RNA Pol and contains the VACV A8 and A23 intermediate transcription factors (A8/A23), only A23 (ΔA8), only A8 (ΔA23), or neither A8 or A23 (ΔA8/ΔA23) were constructed by homologous recombination using GFP and/or DsRed as a reporter. In the scheme shown, for example, A8/A23 virus expresses both A8 and A23, while ΔA8 virus expresses a GFP reporter and A23. (**B**) Expression of A8 and A23 orthologs. BS-C-1 cells were infected with ΔA8 or ΔA23 virus and transfected with plasmids that encode A8 orthologs with V5 epitope tags or A23 orthologs with HA epitope tags regulated by T7 promoters. Following cell lysis, proteins were resolved on SDS-polyacrylamide gels, transferred to membranes, and detected with anti-V5 and anti-HA antibodies. The poxviruses from which the A8 or A23 orthologs were derived are as follows: lanes 1, VACV; lanes 2, MYXV; lanes 3, CNPV; lanes 4, CRV; lanes 5, SGPV. C represents a control plasmid not expressing a protein. The masses in kilodaltons to the left of each blot indicate the mobilities of marker proteins.

For construction of the A8 and A23 expression plasmids, the orthologous ORFs were mammalian-codon optimized to minimize differences in G/C content, which might affect expression. Each plasmid of one set contained an A8 ORF with a V5 epitope tag, while each plasmid of the other set had an A23 ORF with a hemagglutinin (HA) epitope tag. Expression of the proteins following transfection of the plasmids into cells infected with recombinant viruses is shown in **Fig. 3.1B**. Some differences in the intensities of the bands were noted, but the weaker bands were not enhanced by altering the amount of plasmid transfected.

A plasmid with the firefly luciferase (Fluc) gene regulated by the VACV G8 intermediate/late promoter (3, 186) was used to monitor the activities of the intermediate transcription factors. The scheme in which cells are infected with one recombinant VACV of the group expressing T7 RNA polymerase (ΔA8, ΔA23, and ΔA8/ΔA23) and transfected with one or two plasmids of the group containing the T7 promoter (A8, A23, and both A8 and A23) in which A8 and A23 are derived from VACV, MYXV, CNPV, CRV, or SGPV and a second reporter plasmid, G8-Fluc, is outlined in **Fig. 3.2A**. Only the VACV and MYXV A8 genes allowed substantial expression of Fluc (**Fig. 3.2B, left**). In contrast, all of the orthologous A23 genes except the SGPV A23 gene enhanced Fluc in the order VACV > MYXV > CNPV > CRV (Fig. 2B, middle). In the experiments described so far, one of the intermediate transcription factor subunits was from VACV, while the other was one of the orthologs. We considered that the activities of the transcription factor orthologs might be relatively low because they had to work in concert with the VACV factor. Therefore, in the next experiment we infected cells with a VACV lacking both A8

and A23 and transfected two plasmids expressing both transfection factors from the same virus. However, the ranking was similar to that obtained by transfecting A8 or A23 alone (**Fig. 3.2B, right**).

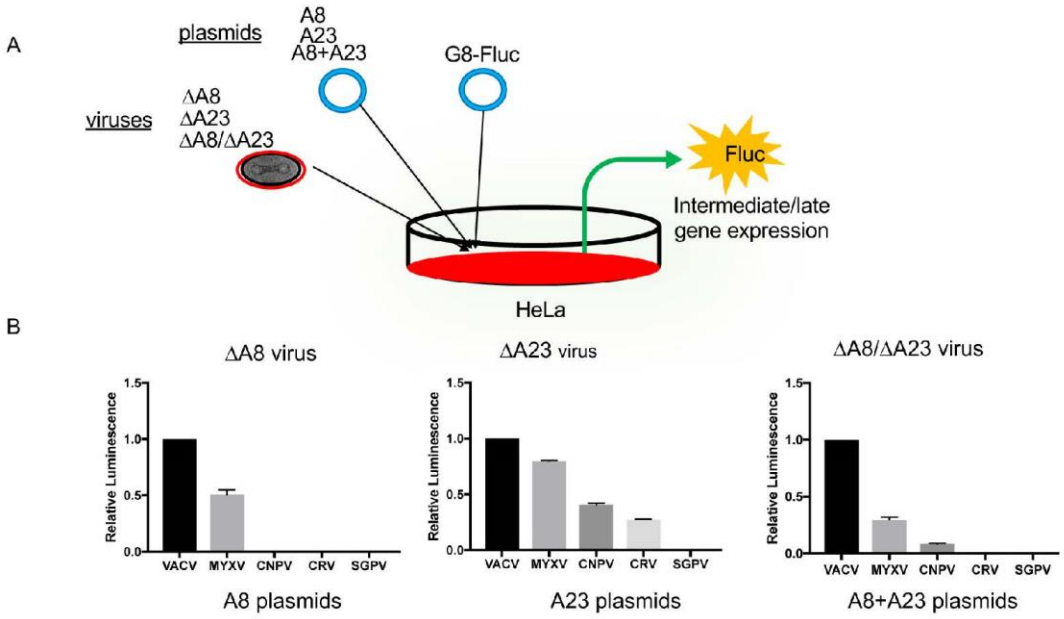

Figure 3.2 Complementation of intermediate gene expression by orthologous transcription factors. (**A**) Experimental plan. Cells were infected with ΔA8, ΔA23, or ΔA8/ΔA23 VACV encoding T7 RNA polymerase and transfected with plasmids encoding A8, A23, or both A8 and A23 (A8_A23) regulated by the T7 promoter and with Fluc regulated by an intermediate/late promoter. Fluc activity depended on expression of A8 and A23 proteins. (**B**) The plan depicted in panel A was carried out by transfecting plasmids into HeLa cells infected with ΔA8, ΔA23, and ΔA8/ΔA23viruses. Relative luminescence values are shown. Experiments were carried out in triplicate, and the bars indicate the standard errors of the means.

3.3.2 Construction of hybrid viruses

We made hybrid viruses in order to further investigate the effects of orthologous transcription factors. Our strategy was to start with viruses that had A8 or A23 replaced by GFP or DsRed and then exchange the fluorescent reporter gene with MYXV or CNPV A8 with an N-terminal V5 tag or with A23 with an N-terminal HA tag by homologous recombination. The recombinant viruses with MYXV genes were detected by formation of nonfluorescent plaques in BS-C-1 cells, and three separate clones with MYXA8 or MYXA23 designated A, B, or C were isolated (**Fig. 3.3A**). The presence of the MYXV genes was confirmed by DNA sequencing. The MYXA8 hybrid made predominantly tiny plaques, whereas the MYXA23 hybrid made intermediate-sized plaques (**Fig. 3.3B**). The different plaque sizes of the hybrid viruses were consistent with the relative activities of MYXA8 and MYXA23 in transfection experiments (**Fig. 3.2B**). A similar attempt to make CNPV and CRV A23 hybrids in BS-C-1 cells failed, evidently because of poor replication.

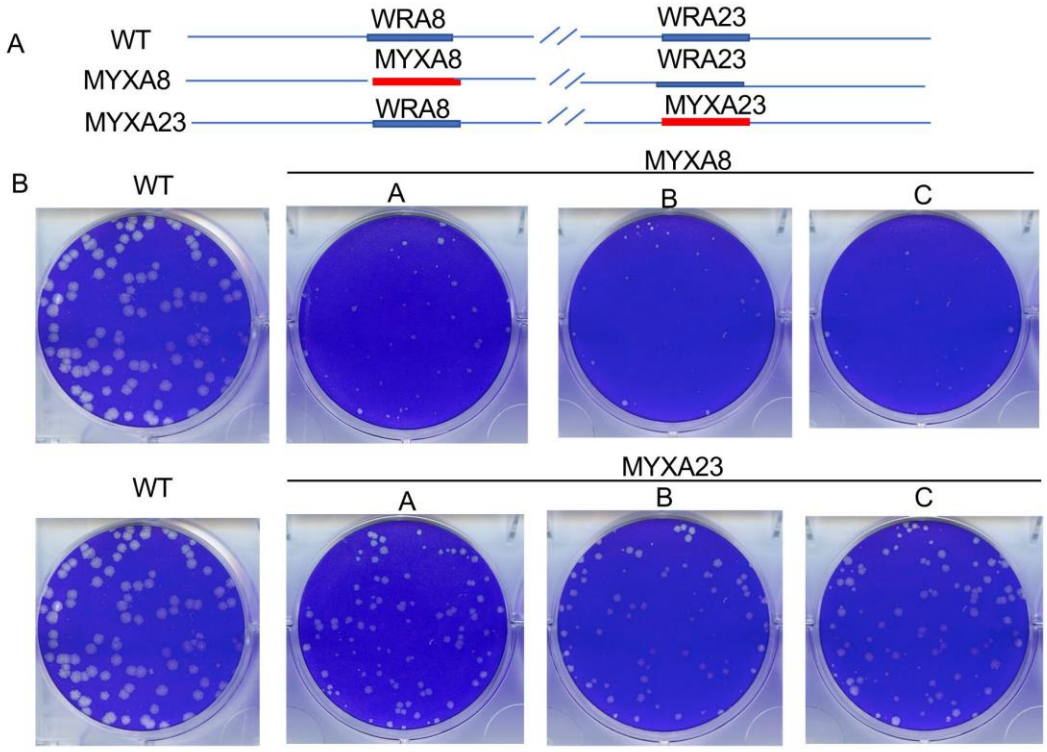

Figure 3.3 Plaque formation by hybrid viruses expressing VACV and MYXV intermediate transcription factors. (**A**) Diagram of viruses expressing the VACV WRA8 and WRA23 (WT), MYXA8 and WRA23 (MYXA8), and WRA8 and MYXA23 (MYXA23). (**B**) Plaques were stained with crystal violet at 48 h after infection of BS-C-1 cells with WT or three independent clones of the hybrid viruses labeled A, B, and C.

3.3.3 Experimental evolution

The inefficient replication of the hybrid viruses containing MYXV transcription factors suggested that beneficial mutations might arise, perhaps in the transcription factors themselves, during serial passaging. The scheme shown in **Fig. 3.4A** was carried out starting with stocks of the three independent plaque-purified clones (A, B, and C) of MYXA8 and MYXA23, which we refer to as passage 0 (P0). Ten passages were carried out, and a portion of each was frozen as the "fossil record." Plaques formed with the MYXA8 virus at passage 10 (P10) exhibited a range of sizes, including many plaques larger than those of P0 (data not shown). In contrast, the plaques formed with MYXA23 did not significantly increase in size after passaging (data not shown). The 24-h yields of the MYXA8 P10 viruses were 3 to 5 times higher than those of the P0 viruses, whereas the yields of MYXA23 did not appreciably change during passage (data not shown) and will not be discussed further.

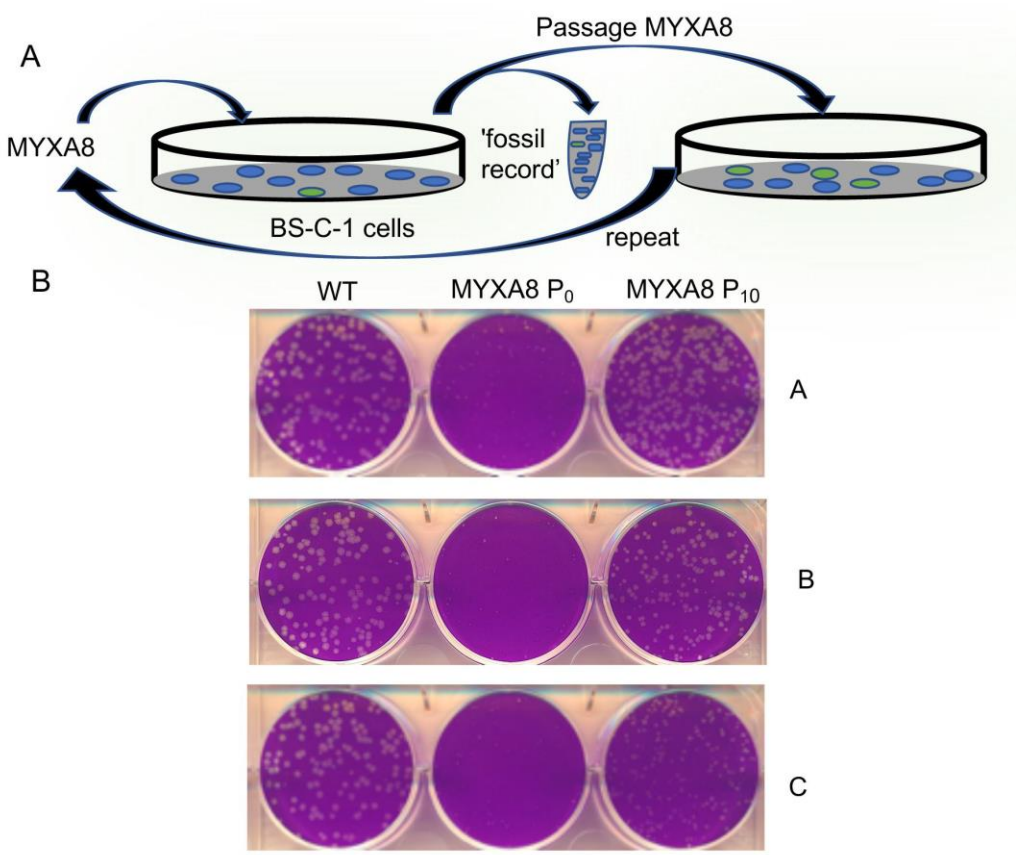

Figure 3.4 Experimental evolution of MYXA8 virus. (**A**) Scheme for serial passaging in BS-C-1 cells of MYXA8 virus containing the MYXV A8 and VACV A23 genes. After each passage, a sample was saved to serve as the fossil record. (**B**) Three clones of MYXA8 virus, **designated A, B, and C in Fig 3.3**, were independently passaged at a multiplicity of infection of approximately 0.1 PFU/cell. Plaques formed in 48 h on BS-C-1 cells by WT virus containing the VACV A8 and A23 genes and by hybrid viruses at passage 0 (MYXA8 P0) and passage 10 (MYXA8 P10) plaque-purified clones from serial passages A, B, and C are shown.

Large plaques were picked from each of the three independent A, B, and C P0 and P10 MYXA8 passages, and the viruses were plaque purified several times. The plaque-purified viruses from the series A, B, and C passage P10 populations are referred to as P10 clones A, B, and C, respectively, to indicate their origin. The plaque sizes of the P10 clones, relative to those of the P0 and the WT virus with VACV A8, are shown in **Fig. 3.4B**. The plaque-purified P10 clones were significantly larger than those of P0, and the P10 A clones approached the size of the WT clones. The enhanced replication of the MYXA8 A, B, and C clones from P10 compared to levels of P0 and the WT are shown in **Fig. 3.5**. At 12 h, the titers of the P10 A and B clones were significantly higher than those of the P0 viruses, and by 24 h all three were higher, with A and B approaching the level of the WT.

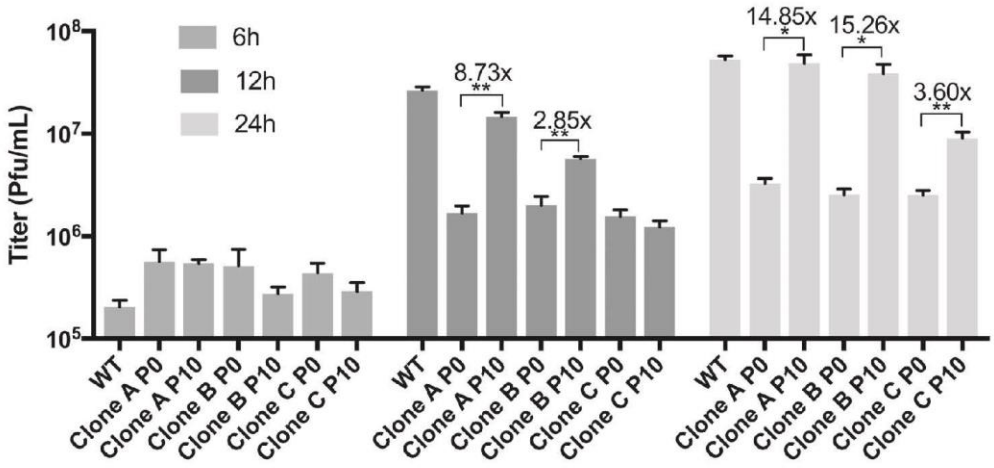

Figure 3.5 Replication of MYXA8 P0 and P10 clones. Triplicate BS-C-1 cell monolayers were infected for 6, 12, and 24 h with 5 PFU/cell of WT virus, and MYXA8 P0 and P10 clones from passages A, B, and C. Virus titers were determined in duplicate by plaque assay in BS-C-1 cells. Bars represent standard errors of the means. *, $P < 0.05$; **, $P < 0.01$.

3.3.4 Genome sequencing

Whole-genome sequencing was performed on the population of viruses in P10 of series A, B, and C and on plaque-purified clones from each series after amplification in BS-C-1 cells. Both synonymous and nonsynonymous changes in the viral genome were found, and the complete sequences have been archived. The genetic changes considered most significant are listed in **Table 3.2**. Single nucleotide nonsynonymous mutations of RNA Pol were detected in each of the passage populations. Two of the mutations are in RPO147 (ORF J6), the largest subunit, and three are in RPO132 (ORF A24), the second-largest subunit. Two of the RPO132 mutations were found in relatively high frequency in the passage B population but were not in the plaque-purified clone from this passage (**Table 3.2**). The five mutations are all different, indicating that they arose independently in the three separate passages. Sequence changes that potentially increase expression of MYXA8 were also found (**Table 3.2**) and will be discussed below. The mutations in the clones were present at 100%, indicating their purity.

Table 3.2 Mutation abundance in MYXA8 P10 passage population and clones

Passage Series	AA change	ORF	Gene	Passage[a]	Clone[b]
A	M424I	J6R	RPO147[e]	30%	100%
	L013M	MYXA8[c]	MYXA8	20%	100%
B	A1212V	J6R	RPO147	6%	100%
	None	-3 MYXA8[d]	MYXA8	18%	100%
	A1026V	A24R	RPO132	14%	0%
	A1133T	A24R	RPO132	30%	0%
C	K1046Q	A24R	RPO132	69%	100%

[a]Passage 10 virus was amplified prior to DNA purification.
[b]Virus from large plaques was clonally purified and amplified.
[c]New translation initiation codon within N-terminal V5 tag preceding MYXA8 ORF.
[d]Improved Kozak sequence consensus sequence preceding N-terminal V5 tag; carries a single nucleotide change at the -3 position relative to the start codon.
[e]RPO, RNA polymerase subunit.

In addition to whole-genome sequencing of the cloned viruses and the P10 population, we performed amplicon sequencing of genomes from sequential passages to determine when changes in the RPO147 and RPO132 sequences occurred (**Table 3.3**). The mutation K1046Q in RPO132 was detected in P2 of the passage series C population and steadily increased in frequency, reaching 55% of the population at P10. The mutation M424I in RPO147 was detected at P4 of the passage series A population and increased to 24% of the population by P10. The mutation A1212V in RPO147 was detected at P8 of the passage series B population and was only 1.9% of the total population at P10. Note that the RNA Pol mutations of passage series A, B, and C increased to 30%, 6%, and 69% of the population, respectively, with the additional passages used to purify DNA for whole-genome sequencing, indicating that selection was still occurring (**Table 3.2**).

Table 3.3 Changes in abundance of RNA pol mutations during passages

	Frequency of mutations (%)[a]		
Passage No.	%RPO147 M424I	%RPO147 A1212V	%RPO132 K1046Q
0	ND	ND	ND
2	ND	ND	1.7
4	3.0	ND	10.1
6	6.1	ND	17.6
8	11.2	1.2	38.1
9	16.8	1.5	50.1
10	24.2	1.9	55.3

[a]ND, not detected.

3.3.5 RNA pol mutations increase replication of the MYXA8 hybrid VACV

The RNA Pol mutations were mapped to the homologous *Saccharomyces cerevisiae* RNA Pol II (RNAP II) subunits **(Fig. 3.6A**). The yeast amino acids corresponding to the mutated ones are located internally, rather than in the solvent-exposed surface where interactions with transcription factors might be expected. To determine the significance of the RNA Pol mutations in the absence of other sequence changes in the hybrid virus genomes, homologous recombination with PCR products containing the mutations was used to directly modify the P0 viruses. DNA segments of approximately 1,000 bp that encompassed the single nucleotide RNA Pol mutations were derived by PCR from the DNA of the cloned P10 viruses and transfected into BS-C-1 cells that were infected with the original MYXA8 P0 viruses. After 20 h, virus was collected, and individual plaques were screened by PCR and Sanger sequencing. Approximately 20 to 30% of the plaques had the desired mutations, attesting to the high rate of recombination, and no additional mutations were found by whole-genome sequencing. Following plaque purification and amplification, the plaque sizes of the mutated and cloned viruses were compared with those of the MYXA8 P0 virus containing the WT RNA Pol sequences. The plaques of the viruses with the RPO147 M424I mutation and the RPO132 K1046Q mutation were larger than those of the original MYXA8 virus (**Fig. 3.6B**), whereas the difference was small in the case of the virus with the RPO147 A1212V mutation (data not shown). Plaque size measurements of all three recombinant viruses are presented in a subsequent figure.

We also determined the virus yields following infection with 0.01 plaque-forming units (PFU)/cell of the MYXA8 P0 clones, recombinant viruses with mutated RNA Pol, and WT virus. Notably, the yields of viruses with the RPO147 M424I mutation from P10 clone A and the RPO132 K1046Q mutation from P10 clone C were higher than those of the P0 MYXA8 A and C clones, which had unmutated RPO147 and RPO132 (**Fig. 3.6C**), indicating that these mutations were adaptive.

A competition experiment was carried out to confirm the ability of the recombinant virus with the RPO147 M424I mutation to enhance replication relative to the level of the original MYXA8 P0 virus. Three BS-C-1 cell monolayers were each infected with 0.01 PFU/cell of a mixture comprised of 90% MYXA8 P0 virus with the WT RPO147 sequence and 10% RPO147 M424I virus for 48 h. The cells were then harvested and titrated, and the virus was passaged two more times. After each passage, PCR and Sanger sequencing were performed to estimate the ratios of the WT RPO147 sequence and the mutated sequence. In the three independent experiments, the percentages of mutated sequence after passage 3 were ~90%, 70%, and 100% **(Fig. 3.6D**), demonstrating the selective advantage of the mutation. In a parallel competition experiment with WT VACV (90%) and WT VACV that had the RPO147 mutation (10%), the relative amounts of the two remained constant during the three passages (data not shown). Thus, the advantage of the RPO147 M424I mutation was specific for the MYXA8 hybrid virus.

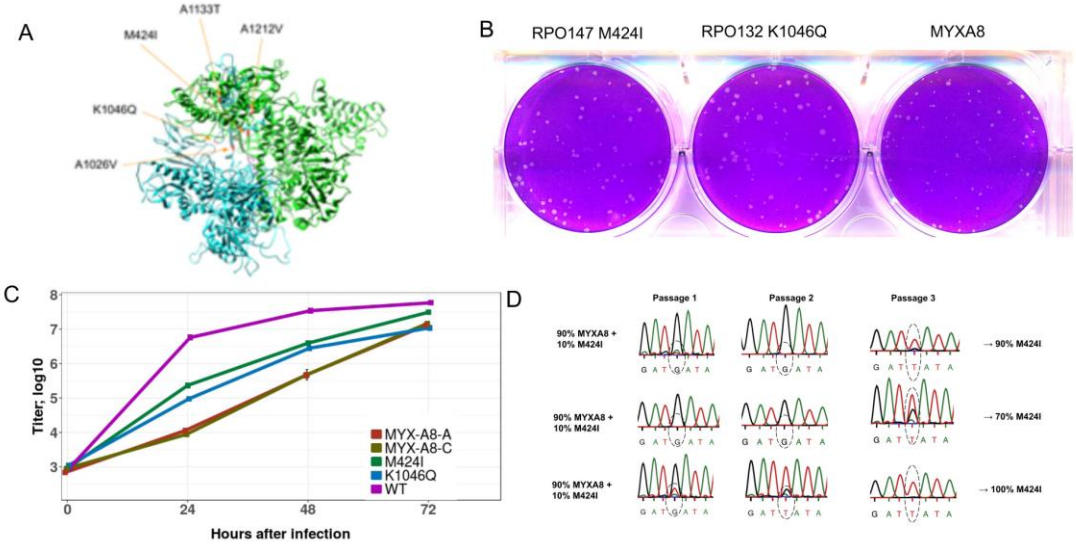

Figure 3.6 RNA Pol mutations. (**A**) Mutations in RPO147 and RPO132 that occurred during passages of MYXA8 were mapped onto the *S. cerevisiae* RNAP II crystal structure. Green, RPB1; blue, RPB2. (**B**) Single nucleotide mutations were inserted by homologous recombination into RPO147 and RPO132 of MYXA8 P0 viruses. Plaques formed by the cloned recombinant viruses and a MYXA8 P0 virus are shown. (**C**) Virus yields at 2, 24, 48, and 72 h after infection with 0.01 PFU/cell of WT, MYXA8 P0 clone A (MYX-A8-A), MYXA8 P0 clone C (MYX-A8-C), and the recombinant viruses with single nucleotide mutations in RPO147 (M424I) or RPO132 (K1046Q). (**D**) Replication competition between the MYXA8 P0 clone A virus and the recombinant virus with mutation in RPO147. BS-C-1 cells were infected in triplicate with 0.01 PFU/cell of a mixture comprised of 90% MYXA8 P0 and 10% MYXA8 with the M424I mutation in RPO147. After each of three rounds of passaging, an aliquot was removed for Sanger sequencing of the region including the M424I mutation. Graphs of the sequence are shown with a dotted oval around the mutated nucleotide. The approximate percentages of M424I after the third passages are indicated at the right.

3.3.6 Mutations of translation initiation sequences of MYXA8 gene

The P10 clones of passage A and B had a mutation that introduced a new translation initiation codon with an optimal Kozak sequence within the N-terminal V5 tag and a mutation that was predicted to improve the original translation initiation sequence of the MYXA8 ORF, respectively (**Fig. 3.7A**). The nucleotides around the translation initiation codon of the V5-tagged MYXA8 in the original P0 clones are CTTATGG, with the ones at -3 and +4 underlined. Although the G at +4 is consistent with good translation initiation, a purine rather than a C is optimal at -3 (195). The single nucleotide mutation in P10 clone A introduced a new translation initiation codon with preferred nucleotides at both -3 and +4 (GGCATGG) within the V5 tag of clone A, while the mutation in P10 clone B changed the C at -3 to a preferred A. Homologous recombination with mutated PCR product was used to introduce these mutations into the MYXA8 P0 clones, similar to the protocol used to make the RNA Pol mutations. Random plaques were selected for Sanger sequencing, and those with the expected mutation were cloned. Comparison of the plaques formed by a MYXA8 P0 clone and the recombinant viruses revealed that the latter produced larger plaques (**Fig. 3.7B**). As anticipated, the V5-tagged MYXA8 protein was expressed more highly in cells infected with the virus with the clone B mutation than with the MYXA8 P0 clone, whereas E3, another early viral protein, was expressed in similar amounts (**Fig. 3.7C**). A comparable Western blot analysis of the P10 clone A mutant was not possible because of the interruption of the V5 tag.

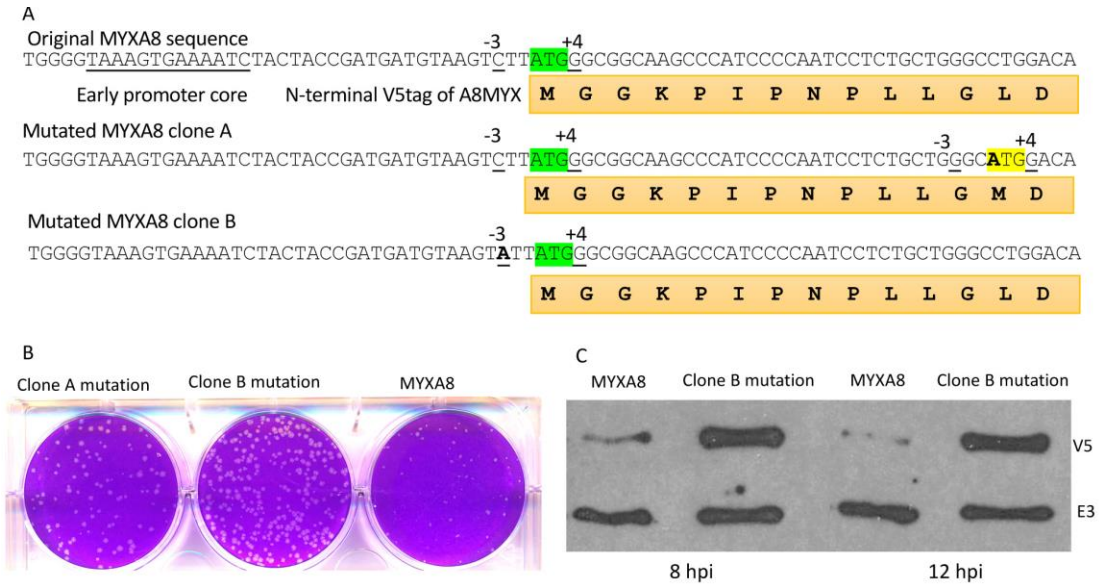

Figure 3.7 Mutations at translation initiation sites of the MYXA8 gene. (**A**) The nucleotide and amino acid sequences around the N-terminal V5 tag of the MYXA8 gene in the original construct and the nucleotide mutations in the P10 A and B clones are shown. The ATG start codon at the start of the tag sequence is shaded green, and the nucleotides at the -3 and +4 positions are underlined. The new ATG within the V5 tag of the clone A mutant is shaded yellow. (**B**) Single nucleotide mutations were inserted by homologous recombination into the MYXA8 P0 virus genome to introduce a new start codon within the V5 tag (clone A mutation) or at the -3 position relative to the original start codon (clone B mutation). Plaques formed in BS-C-1 cells after 48 h by the cloned recombinant viruses and the MYXA8 P0 virus are shown. (**C**) Expression of the MYXA8 protein. BS-C-1 cells were infected with MYXA8 P0 and the virus with the clone B at a multiplicity of infection of 5 PFU/cell. At 8 and 12 hpi, the cells were harvested, and the proteins were analyzed by Western blotting with antibody to V5 to visualize the MYXA8 protein and to E3, another early protein.

3.3.7 Comparison of plaque sizes of clonal isolates and recombinants with single nucleotide changes

The P10 clone A and B viruses had both RNA Pol and translation initiation site mutations. In order to determine the relative contributions of the individual mutations, we compared the plaque sizes of the P10 clones with those of mutants with single nucleotide changes in either an RNA Pol subunit or MYXA8 translation initiation sequence. The measurements in **Fig. 3.8** are presented as box-and-whisker plots: the top and bottom of the box show the upper and lower quartiles, the horizontal line within the box is the median, and the vertical lines are the maximum and minimum values, with the dots representing outliers that are more or less than 3/2 of the upper and lower quartiles, respectively. The plaque sizes of the viruses with individual RNA Pol mutations were each larger than those of the MYXA8 P0 virus. The plaques formed by the K1046Q mutant virus were similar in size to those of the P10 clone C virus, confirming that the latter had no other significant adaptive mutations, as indicated in **Table 3.2**. In contrast, the plaques formed by the M424I and A1212V RNA Pol mutant viruses were smaller than the clone A and B P10 viruses, which also had mutations in the MYXA8 Kozak consensus sequences (**Table 3.2**). Furthermore, the plaques of the mutants with changes in the Kozak consensus sequences were also larger than those of the MYXA8 P0 virus, indicating that the latter as well as the RNA Pol mutations were adaptive. However, none of the P10 clones made plaques as large as those of the WT VACV (**Fig. 3.8**). The color coding in Fig. 8 groups the mutations present in the cloned viruses as detailed in the legend.

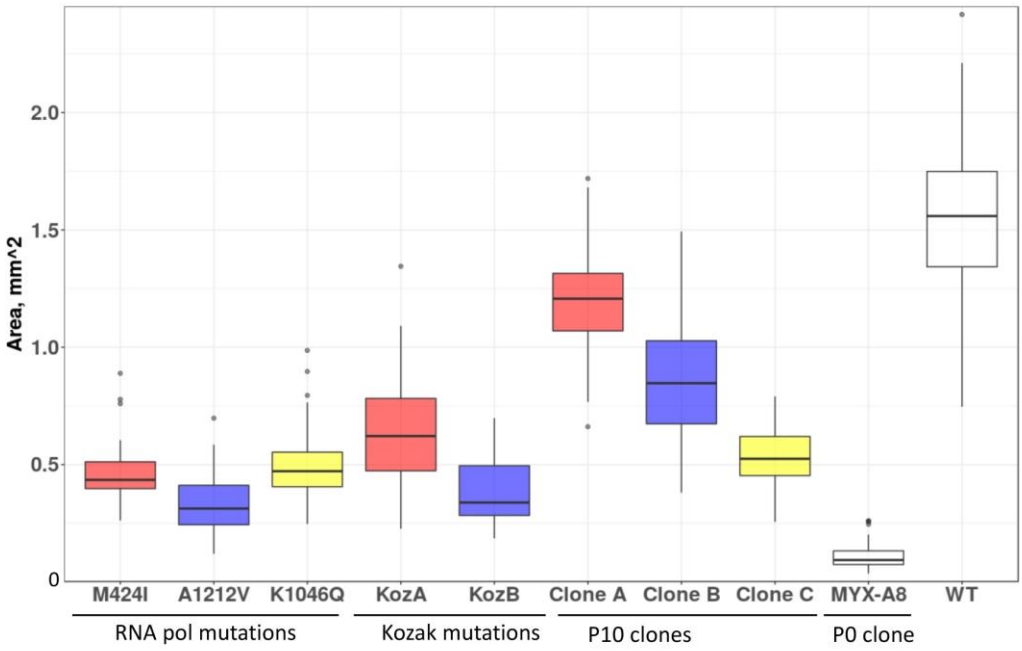

Figure 3.8 Comparison of plaque sizes of recombinant and passaged virus clones. The plaque sizes of recombinant viruses with RNA Pol or Kozak sequence mutations, of MYXA8 P10 clones, of MYXA8 P0 clone A, and of WT virus with the VACV A8 gene are shown. Three independent wells, for a total of 150 to 300 plaques, were analyzed for each virus. Box-and-whisker plots are shown. Color coding: red, clone A and recombinant viruses with individual mutations in that clone; violet, clone B and recombinant viruses with individual mutations in that clone; yellow, clone C and the recombinant virus with the individual mutation in that clone.

3.4 Discussion

Experimental evolution studies are usually carried out with RNA viruses, which are facilitated by high rates of mutation compared to those of DNA viruses. The low mutation rate of VACV has been attributed to intrinsic 3`→ 5` exonuclease activity of the viral DNA polymerase (196-198). Nevertheless, recent studies have shown that adaptation of VACV host range mutants with deletions of genes involved in evasion of the PKR/eIF2α antiviral pathway can occur both by gene amplification and point mutations during serial passaging (69, 146, 174, 185). We initiated the current project to investigate experimental evolution of essential genes and chose the A8 and A23 intermediate transcription factor genes for this purpose. Since there were no mutagenesis studies of these essential proteins, we replaced the VACV factors with orthologs from other chordopoxvirus genera. Transfection experiments indicated that the ability of orthologs to functionally substitute for their VACV counterparts corresponded directly with amino acid sequence identity values. We were able to individually substitute the genes encoding the MYXA8 and MYXA23 proteins, which had ~60% amino acid identity with the VACV proteins, but not those of CNPV and CRV, which had 51% and 40% identity, respectively. Both MYXA8 and MYXA23 hybrid viruses made smaller plaques than the WT virus, but MYXA8 produced the smallest, providing a better model virus for experimental evolution.

Viruses that made larger plaques and replicated to higher titers than the original VACV MYXA8 virus arose during each of three parallel serial passages. Since the hybrid viruses had an MYXV intermediate transcription factor, we had anticipated that adaptive mutations might occur in either the MYXA8 or VACVA23

intermediate transcription factor subunit or in the RNA Pol. Although we found no mutations in either transcription factor subunit, two different single nucleotide mutations in two independent passage series predicted more efficient translation of MYXA8 mRNA. More interesting was the occurrence of RNA Pol mutations in each of the three parallel passages: there were two independent mutations in RPO147 and three in RPO132, the largest and second largest of the eight subunits in VACV RNA Pol, respectively. The RNA Pol mutations were not detected in P0 but varied from 6 to 69% of the entire population after P10, suggesting strong selection. Amplicon sequencing of specific RNA Pol mutations confirmed progressive increases in mutation frequency in successive passages. Although we did not detect gene duplications in the P10 populations, such changes could have occurred transiently in earlier passages that were not examined by whole-genome sequencing.

Three MYXA8 viruses were clonally purified from the P10 populations, each of which formed larger plaques than the MYXA8 P0 clones. The P10 A and B clones each had mutations in RNA Pol and the translation site of the MYXA8 gene, while P10 clone C had only an RNA Pol mutation. To attribute phenotypic changes to single nucleotides, we mutated the RNA Pol of MYXA8 P0 viruses by transfection of ~1,000-bp PCR products. The high rate of recombination, which allowed this simple procedure without a selection marker, is not generally appreciated and for this purpose is far simpler than the transient dominant selection method (199). The absence of any other changes in the mutant viruses was verified by whole-genome sequencing. These recombinant viruses allowed us to prove that RPO147 M424I and RPO132 K1046Q increased plaque size and virus yield substantially, although the

RPO147 A1212V mutation had a smaller effect. Furthermore, a competition experiment carried out between the RPO147 M424I mutant and the MYXA8 P0 with WT RPO147 virus demonstrated that the mutant rapidly became dominant. The plaque size and virus yield of the RPO132 K1046Q mutant were similar to those of the P10 clone C virus, confirming the absence of other adaptive mutations in the latter. While the RPO147 M424I and RPO147 A1212V mutant viruses made larger plaques than the MYXA8 P0 clones, they were smaller than those of the P10 A and B clones, suggesting that the MYXA8 translation site mutations in the P10 clones were responsible. The MYXA8 gene in the starting virus has a C at the -3 position and a G in the +4 position relative to the first nucleotide of the translation codon. According to the Kozak model, G at +4 is a good context for high translation initiation, but a purine is preferred to the C at -3 (195). Interestingly, a C→A mutation in the -3 position relative to the ATG start codon of MYXA8 occurred in one P10 clone. In the other P10 clone, the point mutation created a new ATG with a good Kozak consensus sequence within the N-terminal V5 epitope tag. Genome sequencing of the passage 10 populations revealed a frequency of 18% for the C→A mutation and a frequency of 20% for the new ATG, suggesting that selection had occurred. Homologous recombination was used to make these exact mutations upstream of the MYXA8 open reading frame of the original virus. In each case, the mutants enhanced virus replication, as shown by the presence of increased plaque size. In addition, Western blotting with anti-V5 antibody showed that the C→A mutation increased expression of MYXA8. Although the mutation within the V5 tag likely increased expression as well, this could not be confirmed because of the inability of anti-V5 antibody to

detect the protein. Curiously, the Kozak mutation within the V5 tag conferred higher replication than the upstream Kozak mutation, suggesting the possibility that the V5 tag is deleterious and that the disruption is beneficial. However, previous studies showed that addition of a V5 tag to the VACV A8 protein did not have a noticeable effect (118).

Based on homology with the yeast RNA Pol, the mutations in RPO147 and RPO132 were predicted to be located internally near the catalytic site rather than in a solvent-accessible region where interaction with other proteins might occur. This would suggest that the mutations do not directly enhance binding of the transcription factors. Some mutations found by other investigators in RPO147 (A535V and S288Y) and RPO132 (Y462H) confer resistance to isatin-β-thiosemicarbazone (IBT) by reducing transcription elongation (121, 122). In addition, mutants in RPO132 were found during experimental evolution with E3 deletion mutants that have defects related to doublestranded RNA with activation of the oligoadenylate synthetase (OAS)/RNase L and PKR/eIF2α pathways. In one report, an RPO132 T1120M mutation partially reduced PKR phosphorylation and prevented eIF2α phosphorylation, resulting in increased viral protein synthesis and resistance to IBT (10), whereas in another report, an RPO132 Leu18Phe mutation increased levels of PKR and eIF2α phosphorylation; an RPO132 Lys452Asn mutation elicited no change in PKR and eIF2α phosphorylation compared to levels in the E3 deletion mutant (174). Thus, the previously described RNA Pol mutations appear to have diverse effects.

We plan to investigate the role of RNA Pol mutations in enhancing replication of MYXA8 in a follow-up study. We are considering several possibilities: (i) that the mutated RNA Pol is more promiscuous and less dependent on intermediate transcription factors for RNA synthesis, (ii) that the mutations slow down transcription, allowing more time for the heterologous transcription factors to act, and (iii) that diminished processivity may reduce double-stranded RNA and indirectly enhance replication. With regard to the last idea, it will be interesting to determine the effects on replication of MYXA8 of previously described mutations in RPO147 and RPO132. To conclude, we have shown that accelerated experimental evolution can be achieved by swapping orthologous poxvirus transcription factor genes and suggest that this approach is generally applicable to interrogating additional gene functions.

3.5 Materials and Methods

3.5.1 Construction of deletion and hybrid viruses

RK13 A8-23 cells (118) in 24-well dishes were infected with 0.05 PFU/cell of vTF7-3, a WR strain of VACV that encodes the bacteriophage T7 RNA Pol regulated by a VACV early/late promoter (126). After 1 h, the cells were washed with fresh medium and transfected with 2 μl of Lipofectamine 2000 (Invitrogen) and 500 ng of a PCR product encoding GFP or DsRed regulated by the VACV p11 late promoter and sequences flanking the A8 or A23 gene, respectively. In the case of the A8 deletion, all of the N-terminal nucleotides of as well as three nucleotides in the preceding intergenic region were deleted, but 28 nucleotides of the C terminus were retained because of overlap of the A9 gene. In the case of the A23 gene, all N-terminal

nucleotides were deleted, but 47 nucleotides at the C terminus were retained because of overlap of the A24 gene. The medium was replaced after 24 h, and the incubation continued for 48 h. Diluted cell extracts were used to infect fresh RK13 A8-23 cell monolayers, which were then overlaid with 0.5% methylcellulose. After 48 h, fluorescent plaques were picked with sterile toothpicks. The viruses were clonally purified by repeated plaquing on RK13 A8-23 cells, and the purity was assessed by PCR. Deletion of both A8 and A23 from the same virus was carried out by successive deletion of the individual genes.

Hybrid viruses containing orthologous mammalian codon-optimized A8 or A23 genes were constructed using a similar recombination protocol in HeLa cells, except that the mammalian-optimized ortholog replaced GFP or DsRed, and identified by formation of non-fluorescent plaques in BS-C-1 cells.

3.5.2 Construction of expression plasmids

Nucleotide sequences for the A8 and A23 orthologs were obtained from GenBank and included VACV WR (accession number NC_006998), MYXV-6918 (EU552530), CNPV-VR111 (NC_005309), CRV (NC_008030), and SGPV (NC_027707). The genes were codon optimized for expression in mammalian cells and synthesized by Life Technologies GeneArt Gene Synthesis service. Primers were designed to amplify genes from the plasmids supplied by the company using Phusion High-Fidelity (HF) master mix with HF buffer (M0531L; New England Biolabs [NEB]). The PCR products were purified using a QiaQuick PCR purification kit (catalog no. 28104) and inserted into a Blunt TOPO vector (45-0245; Invitrogen). A8 and A23 ORFs were tagged at their N termini with V5 and HA, respectively. Plasmid

DNA was extracted using a QiaPrep Spin Miniprep kit (catalog no. 27106). Colonies of transformed Escherichia coli were picked and grown, and the DNA was sequenced.

3.5.3 Western blotting for analysis of plasmid expression

HeLa cells were infected with 3 PFU/cell of vTF7-3 for 1 h and then transfected with 500 ng of plasmid in Lipofectamine 2000. Samples were incubated overnight at 37°C and harvested. Anti-V5 (ab27671; Abcam) and anti-HA (H3663; Sigma) antibodies were used to detect expression of the viral proteins. Secondary antibodies were fluorescently labelled (1:10,000; Li-Cor). Bands were visualized using a Li-Cor Odyssey CLx.

3.5.4 Luciferase assays

HeLa cells were infected in triplicate with 3 PFU/cell of the deletion viruses and then transfected with 500 ng of one or both A8 and A23 plasmids and the luciferase plasmid. Cells were incubated at 37°C for 12 h, and luciferase activity was measured (catalog no. E1501; Promega) with a luminometer (Berthold Detection Systems) and normalized to the activity of the WR orthologs. Samples were diluted to keep readings in the linear range.

3.5.5 Serial passaging of hybrid viruses

BS-C-1 cells in 10-cm-diameter dishes were infected with 0.1 PFU/cell at 37°C, harvested at 48 h, and then resuspended in 1 ml of medium. The cells were frozen and thawed three times and then sonicated before each passage.

3.5.6 Whole genome sequencing

DNA was purified as described by Esposito et al. (200) with minor modifications. T75 flasks of BS-C-1 cells (~$1x10^7$ cells) were infected and 2 to 3 days later collected by scraping. Cell pellets were washed with phosphate-buffered saline (PBS) and resuspended in 400 µl of 20 mM Tris, pH 8.0, 5 mM EDTA, and 1% Triton X-100; 10 min later, NaCl was added to a final concentration of 0.2 M. The cell extract was centrifuged at 800 x g for 3 min, and the nuclear pellet was discarded. The supernatant was treated with proteinase K in 0.5% SDS for 1 h at 50°C, and the DNA was extracted with phenol. Two volumes of ethanol were added to the aqueous phase; the visible threads of DNA were collected by brief centrifugation and washed twice with 80% ethanol, and the pellet was dissolved in 100 µl of water. Genome sequencing was done with an Illumina Miseq 2- by 150-bp configuration by the Genewiz company or the NIAID core facility.

3.5.7 Amplicon sequencing

Amplicon sequencing was carried out by either Illumina or Sanger sequencing. For the latter, the PCR primers were designed to produce a product of about 400 bp with the mutation in the middle. Cells were lysed by three freeze/thaw cycles and sonication, and 4 µl of the lysate was used for PCR in 50-µl reaction volumes (Phusion High-Fidelity PCR Mix; NEB). PCR products were purified by a Qiagen PCR kit, and Sanger sequencing was performed with the original primers used for PCR. The percentage of a mutation was estimated based on the heights of the peaks corresponding to specific nucleotides. For better quantitation, primers were designed to produce products of about 150 bp with the mutation in the middle, and

the PCR products were analyzed by ultradeep sequencing on the Illumina Miseq platform by Genewiz.

3.5.8 Introduction of single nucleotide mutations into the VACV genome

BS-C-1 cells were infected with 3 PFU/cell of VACV and after 1.5 h transfected with a PCR product of approximately 1,000 bp with the desired nucleotide change in the middle. At 18 to 20 h after infection, the cells were lysed, and diluted virus was used to infect BS-C-1 cells. At 48 h, the virus from individual plaques was placed in 250 µl of medium and subjected to three freeze-thaw cycles, and 4 µl of the lysate was used for PCR with the original primers in a 50-µl reaction volume (Phusion High-Fidelity PCR Mix; NEB). PCR products were purified with a Qiagen PCR kit, and clones with mutations were identified by Sanger sequencing. The clones with the desired mutations were plaque purified once more; 100 µl of the virus from the plaque was used to infect wells of a 24-well plate of BS-C-1 cells, and virus was collected after 2 to 3 days when the cytopathic effect was complete. For the second passage, six wells of a six-well plate were each infected with aliquots from the first passage.

3.5.9 Plaque size determination

BS-C-1 cells were infected and covered with 0.5% methylcellulose. After 48 h at 37°C, the plates were stained with 0.1% crystal violet and scanned using a Canon 9000F Mark II color image scanner. Plaque sizes were determined using ImageJ, version 1.51w, software. Three independent wells, for a total of 150 to 300 plaques, were analyzed for each virus.

3.5.10 Statistical analysis

Statistical analysis was carried out using GraphPad Prism (version 7.0c). Unpaired two-tailed t tests were performed and P values of <0.05 were deemed statistically significant.

3.5.11 Competition experiment

Approximately $2.5x10^5$ BS-C-1 cells were infected in triplicate with 0.01 PFU/cell of a virus mixture comprised of 90% MYXA8 P0 clone and 10% MYXA8 with the M424I mutation in the RPO147 gene. After 48 h, the cells were resuspended in 250 µl of medium. Virus was released by three freeze-thaw cycles and sonication, which was followed by two additional passages in BS-C-1 cells at 0.01 PFU/cell for 48 h.

3.5.12 Mapping of VACV RNA pol mutations on the eukaryotic RNA pol structure

Alignments between the VACV J6 (RPO147) and A24 (RPO132) gene products and *S. cerevisiae* RPB1 and RBP2 subunits of RNAP II were generated using the PRALINE web server (201) (http://www.ibi.vu.nl/programs/pralinewww/). Corresponding J6 and A24 variant residues were then mapped onto the *S. cerevisiae* RNAP II crystal structure (PDB accession number 5U5Q) by use of Chimera software (202) (http://www.cgl.ucsf.edu/chimera/). Accession number(s). Data are available at the Sequence Read Archive under BioProject accession number PRJNA481995.

Chapter 4: Discussion and Future Directions

4.1 Functional Compatibilities and Incompatibilities of Poxvirus Intermediate Transcription Factor Orthologs

It was established that the promoters between different genera of POXVs were functionally interchangeable (36-38, 194). In this project, we wanted to determine if this applied to the intermediate transcription factors. We chose the intermediate transcription factors because little is known about them and both deletion viruses and a complementing cell line had previously been generated in our lab by Warren et al. (118). The viruses used in this study were generated in a similar fashion except that instead of the VACV WR backbone being used, the vTF7-3 parental virus was used. This virus, also generated in our lab by Fuerst et al. (27, 126), constitutively expresses the T7 RNA Polymerase. Our plan was to put the ITF orthologs used into a plasmid under a T7 promoter, so they could be expressed when transfected in with the knockout viruses. The next step was choosing the orthologs to test. We did not work with the closely related ones, such as CPXV and MPXV, since those were likely to work just as well as the native wild-type genes, which were going to be used as positive controls. Our initial plan included using orthologs from Sheeppox (SPV) but biosecurity/biosafety concerns were raised by the company performing the codon optimization. We then chose from a list of safer options and thus settled on our current list of orthologs which we felt would provide varying levels of rescue, including one which we thought might not work under the experimental conditions, the newly discovered SGPV (164). Once the deletion viruses and the epitope-tagged

ortholog plasmids were generated, we had to check that they all behaved as expected before proceeding with the initial experiments.

The results of the luciferase assays turned out as we had expected, with the orthologs giving a spectrum of activities, with the SGPV orthologs having no activity. Even more fittingly, the level of complementation corresponded to the level of amino acid identity, relative to the WR genes, even when both subunits came from the same virus. This meant that there were inherent differences between the orthologs. Building off these results, we then wanted to see if we could generate replication-competent viruses with the compatible orthologs. We were able to do so successfully with the two MYXV orthologs but not the CNPV and CRV orthologs, perhaps due to reduced compatibility with the other components of VACV transcription system. We then reasoned that, if passaged enough times, these hybrid viruses would accumulate mutations that increase viral replication. We thought that the most likely candidates to mutate would be the orthologs themselves, followed by the vRNAP. During and after the experiment, we observed noticeable increases in viral fitness associated with the hybrid virus containing the MYXV A8 ortholog. WGS results confirmed the presence of mutations not present in the parental or starting viruses. While we did discover mutations present in the two largest subunits of the vRNAP, we did not find any mutations within the MYXV A8 ortholog. Additionally, those mutations that were in the vRNAP all seemed to map to the interior of the enzyme, making it unlikely they were involved in interactions with the ITFs. These findings led us to speculate that these mutations may affect the processivity or promiscuity of the enzyme instead.

4.2 Characterization of Mutations

Our initial focus was on the mutations found in the vRNAP. We found it particularly interesting that independent passages had their own unique RNAP mutations with a total of 5: 2 in the largest subunit and 3 in the second largest subunit. We had reasoned that the vRNAP would be the second likeliest place for adaptive mutations to arise. Upon re-examining our sequencing results, however, we also found two independent mutations in upstream regions of the MYXV A8 ortholog in two of the three independent passages. The nature of these two mutations, which improved the Kozak consensus sequence, led us to think that they might result in increased expression of the MYXV A8 ortholog, which should be of benefit to the virus. Our first step, though, was to confirm that the RNAP mutations increased in frequency in their respective populations over the course of the experiment, which would be expected if they were indeed beneficial. This was done using both Sanger and Amplicon sequencing, which supported the idea that these mutations increased viral fitness. This still needed to be confirmed experimentally though, so each mutation was introduced individually into the starting virus via homologous recombination. Each mutation on its own enhanced the ability of the virus to replicate, as measured by plaque and spread assays. The mutations found related to the MYXV A8 ortholog expression were reassuring in two regards. First, we had thought this would be the most likely place for mutations to arise, although we hypothesized they would be within the gene itself. Secondly, they helped explain why our initial experiments introducing the RNAP mutations back into the starting virus did not fully reproduce the enhanced-growth phenotypes observed in the

passage-derived mutants. These mutations only appeared to be beneficial in the context of the poorly-replicating hybrid starting virus, and not the wild-type virus, in the competition assay, at least regarding RNAP Mutation 1 or A.

Our next step in the process of characterizing these mutations was trying to determine the mechanism(s) by which they enhanced viral replication and if there were any fitness trade-offs associated with them. The A24R gene has been implicated in experimental evolution studies before as a prime target in which adaptive mutations arise (69, 174, 175). We wanted to see if our hypothesis regarding increased expression of the MYXV A8 ortholog would be supported by data, so western blots were done to look at levels of expression compared to the viruses with no mutations. In the cases of P10B, P10C, MutB, expression of the MYXV A8 ortholog was greater at certain time points compared to viruses without the mutation present, including the starting virus (see **Fig. A3**). This is obviously beneficial to the virus, but how is still uncertain. Unfortunately, we cannot say this is the case with P10A/MutA right now, as the V5 epitope tag is no longer expressed and thus cannot be recognized by the antibody. D13 protein expression was not significantly different between WT, the P0 clones, and the P10 clones (see **Figs. A2 and A3**). Assays using ddPCR were also performed with the recombinant viruses, but the results did not show any significant differences in transcript levels between them and the P0 virus (see **Fig. A4**).

IBT resistance results showed a range of phenotypes (see **Fig. A5**). The IBT-resistant mutant first characterized by Condit et al. (120) was used as a positive control against which other RNAP mutations were tested, while wild-type virus and

the P0 virus were used as negative controls. What we found was that certain mutations, specifically MutC and the B Mut, conferred more resistance than MutA in comparison to the wt and P0 viruses. The resistant mutant, IBT^{r90}, grew the best in the presence of the drug. The titer of the virus was only about a log or so higher than the titers for the MutC and B Mut (see **Table A1**) viruses. MutA has mutation in RPO147, while MutC is an RPO132 mutation, just like IBT^{r90}. What these results tell us though is that even though these mutations all provide a growth advantage, they may do so using different mechanisms and with different fitness tradeoffs. A plaque size reduction assay was also attempted but it was unsuccessful.

Western blots were also done to see if PKR and eIF2α phosphorylation patterns/levels were affected compared to wild-type, but the results were inconclusive. They would also have to be done and/or repeated with the recombinant viruses. Various mutations in A24R have also been implicated in the PKR/eIF2α pathway, with some found to increase activation while others had no effect, even though the mutations were all beneficial to the virus (69, 174). This suggests a more complicated mechanism in which these mutations affect viral growth in different ways despite resulting in the same overall phenotype. Theoretically, IBT resistance and the PKR/eIF2α could be connected, as those IBT-resistant mutants may be defective in transcription elongation, as was found with the Condit mutation (121). This should in turn to the less dsRNA being produced, which should reduce the amount of PKR activation. As such, certain phenotypes could be tied together conceivably. In this case, the more IBT-resistant the mutant is, the lower amount of PKR activation will be seen.

4.3 Future Directions

There are numerous avenues that could be explored on this project. One would be to attempt again to generate the hybrid viruses with the CNPV and CRV orthologs, if it is indeed possible. It would be interesting to characterize those viruses and then passage them similar to the MYXV hybrid viruses. Would the same or similar mutations arise, or would we discover more drastic changes in the viral genome since neither of these orthologs worked as well as either of the MYXV orthologs? Alternatively, it would be interesting to see if the mutations that enhanced MYXA8 replication would also enhance CNPV and CRV.

The mechanism by which these various mutations benefit the virus still need to be determined. Is there increased transcription or translation? Is it specific to intermediate and late genes or are early genes affected? This would be particularly interesting regarding the RNAP mutations, as we would be able to state whether they have a global effect or only affect certain classes of genes. As for the transcriptional regulation, the ddPCR experiments would need to be repeated to get clearer results, but the answer seems to be no as of right now. This appears to also hold true for the translational regulation. Another possible way to examine transcription would be to do an *in vitro* assay. This technique, however, requires a lot of purified materials and is more labor-intensive than ddPCR.

While it seems these RNAP mutations do not benefit wild-type virus, it would be ideal if we could also test the A8 mutations but that is impractical in the case of Mutation A. Perhaps mutation B could be tested, but we do not have antibodies against the native proteins, which is why the codon-optimized versions were epitope-

tagged. This problem could be solved if a revertant virus was made using the codon-optimized and tagged VACV A8R gene. This way we could actually measure expression of the VACV A8 protein and see if the mutation increases said expression, which we would expect it to. Granted, this is not the same as looking at the native protein in the wild-type virus, but this would be the best we could do until antibodies against the native protein were generated. Unfortunately, the nature of Mutation A rendered the V5 tag irrelevant, so as of right now we cannot test whether expression of the MYXV A8 ortholog is indeed increased in this case, as with Mutation B. The secondary start codon would also negate the workaround mentioned above with the hypothetical VACV A8 revertant as the V5 tag would still not be within the coding region. Checking expression with Mutation A would require an antibody to the native A8 protein itself or the codon-optimized version if it was attempted with a revertant. The nature of Mutation A, however, may make it unnecessary in the wild-type virus, as a secondary start codon would not be needed, and perhaps even be deleterious.

All the experiments described relating to the blind passaging and the isolation and characterization of these various mutants were carried out in the same cell line, BS-C-1 cells. Given this fact, it would be interesting to see if these mutations still provide a benefit to the virus outside of this cell line. Such a phenomenon would not be unprecedented in this context, as this was observed by Cone et al. (174). One of their A24R mutations, Lys452Asn, provided a benefit that was not species or cell type specific, whereas their other A24R mutation, Leu18Phe, did. As such, it may be prudent to test both the isolated P10 mutants and the recombinant mutants in other

cell lines to see if plaque formation, spread, protein expression, and/or viral growth/yield are significantly affected. Some of these other cell lines could include HeLa, A549, RK-13, and BHK 21. While it would be easier if all the mutations provided a universal benefit, regardless of cell line, it would much more interesting if differences, especially significant ones, were seen between the different cell lines. The challenge would then be to determine why certain mutations are more beneficial in certain cell lines or why certain cell lines are able to, or better able to, negate the enhancement in replication provided by these mutations.

Another aspect of the RNAP mutations that needs to be tested is the question of IBT resistance, which has previously been mapped to the A24R gene (69, 120-122). In the case of Brenan et al., the A24R mutation which conferred resistance to the drug also happened to be an adaptive mutation that arose during experimental evolution (69). Other IBT-resistant mutants have been mapped to the J6R gene (122). IBT seems to act on the process of either transcription elongation or termination, as it causes read-through transcription leading to the production of transcripts longer than those found in wild-type virus. These read-through transcripts are more likely to cause the accumulation of excess dsRNA, which activates the cellular 2-5A pathway and induces RNaseL activity, inhibiting viral growth (121). J6R and A24R encode RPO147 and RPO132, respectively, the two largest subunits of the vRNAP, so it is not surprising that these genes would be involved in potentially counteracting the effects of this anti-POXV drug. Since all the RNAP mutations found were in one of these two subunits, all should be tested to see if they provide any benefit in the presence of IBT. This could be done by doing plaque and/or spread assays in the

presence of the drug and by testing both the P10-isolated mutants and the recombinant mutants. It would be interesting to see how many provide a benefit. It would also be interesting to see how MutB performed in this assay since it was not initially included. If there are different phenotypes attributed to these mutations, it may then be worth investigating to determine the reason(s) for the differences. It may also be of interest to compare these RNAP mutations to the other that have been previously characterized (69, 120-122, 174), especially those which were not tested for IBT resistance at the time of their initial characterization (174). So far, only the mutation found by Condit et al. (120, 121) has been used for comparison. All these mutations could then be mapped to the crystal structure of the yeast RNAP, which is homologous to the VACV RNAP and whose structure is used in place of the latter since the structure of the VACV one has not been solved yet. A pattern or patterns may emerge regarding the phenotype observed in the presence of IBT and where these mutations are located and/or the nature of the mutations themselves. Regardless, the results could potentially provide insights into the mechanism(s) of the growth benefit provided by these RNAP mutations. The MYXV A8 mutations might have no effect in the context of IBT resistance but should also be tested so that no interesting results are missed.

As mentioned previously, adaptive A24R mutations have been implicated in PKR and/or eIF2α phosphorylation (69, 174), which had not been reported prior to the study by Brennan et al. (69). The RNAP mutations isolated in this study should be examined regarding their effects on this antiviral pathway, as has been done with other similar mutations. This should be done for both the P10-isolated and the

recombinant mutants. As stated above, previously-characterized A24R mutations do not have a straightforward relationship with the PKR/eIF2α pathway (69, 174), so it will be fascinating to see if the mutations found in this study also have a range of phenotypes. Then, as with the IBT characterization, it may be of note to map these mutations to the RNAP crystal structure and see if any patterns emerge. As with the IBT screening, the MYXV A8 mutations should also be investigated, if for no other reason than to confirm that they do not affect this antiviral pathway, although it is possible there may be an effect.

Another avenue that can be explored is to isolate additional adaptive mutations, either through additional passaging of the blindly-passaged viruses used here or new rounds of independent blind passaging of the starting viruses used, both of which have already been done (data not shown). These have led to the identification of an additional mutation in the vRNAP (B Mut), another Kozak mutation, and mutations in other genes, such as G6, A26, and F10 (see **Table A1**). All seem to be beneficial, though the specific mechanisms are not yet known. Two additional mutations were also found after virus amplification when samples were being prepared for WGS, both of which were also in the RNAP (see **Table 3.2 and Fig 3.6A**). Some of these new mutations have already been introduced to generate recombinant viruses, with preliminary results showing that they are indeed beneficial (data not shown). It would be prudent, however, to take a step back and characterize the library of mutants we currently have now instead of continuing to identify and isolate more and accumulating adaptive mutations for which we do not know the mechanisms.

Lastly, it may be worth revisiting the initial luciferase assay experiments to determine if the RNAP mutations or mutations equivalent to the MYXV A8 translation-enhancing ones increase transcriptional activity of all or any of the orthologs used. Along this same line of thinking, perhaps these mutations could provide some benefit to the MYXV A23 hybrid viruses, which already grew well enough that no obvious adaptive mutations were found after blind passaging. Perhaps then these viruses would grow just as well as wild-type virus instead of almost as well. With regard to the initial goal of the study to investigate the structural composition of the transcriptional complex, perhaps IP/co-IP experiments could be done with these mutants to see if they have any effect on any potential binding affinities between the vRNAP (perhaps using the tagged version of this enzyme used in our lab by Katsafanas and Moss (1999) (203)), the ITFs, and/or the viral DNA, which could provide a bit more insight into the make-up of this complex and the roles of the ITFs.

Appendix A: Characterization of RPO and MYXA8 Mutants

A1 Introduction

Once these mutations were identified and confirmed to be beneficial for the virus, via plaque and spread assays, we wanted to start trying to determine how they provided their benefits. By introducing these mutations into viruses separately, we could isolate the effects and determine that a specific mutation was responsible for a particular phenotype, without the possible confounding effects of the other beneficial mutations. An additional spread assay was performed to show the improvement in viral fitness for the P10-isolated viruses over the respective P0 viruses. One of the first ways we sought to identify the benefit these mutations provided was gene expression. The most obvious way these mutations would act would be increased gene expression at either the translational and/or the transcriptional level. We then wanted to see if these mutations had any other effects, such as conferring resistance to IBT. For the following experiments, the P10 Clones/Mut 1/2/3 refer to the plaque-purified viruses that were picked from the P10 virus population following blind passaging. Viruses referred to as Mutant1 or MutantA refer to the recombinant viruses generated by introducing individual mutations into the P0 starting virus via homologous recombination. It should also be noted that these experiments were only done once. As such, the results need to be confirmed by performing repeats of the respective assay.

A2 Results

A2.1 Spread Assay

A2.1.1 P0 and P10-isolated Viruses

Once these P10 clones were isolated, we wanted to demonstrate that they grew significantly better than the starting viruses. Even though viruses with the individual mutations had not been isolated yet, the mutations had been identified and had been found in these isolates. If these mutations did indeed confer better growth, we would see spread closer to wild-type. BS-C-1 cells were infected at an MOI of 0.01, in triplicate, and samples were harvested at 48 hpi. Viruses were also titrated on BS-C-1 cells. The P10 isolates all grew much better than the P0 viruses, rivalling the titer of wild-type (**Fig A1**). P10A had a titer that was actually slightly higher than that of wild-type. This showed that the P10 clones had significant growth advantages over the starting viruses, demonstrating the blind passaging had led to the generation of beneficial mutations. To help further confirm that the RNAP mutations were at least partially responsible for increased viral fitness, the individual mutations were introduced into the P0 viruses and the spread assay was repeated. The recombinant viruses grew better than the P0 viruses but not as well as the wild-type virus (see **Fig 3.6C**). This was because these viruses only contained a single beneficial mutation, rather than the two found in the case of P10A. These results also supported what was seen with the recombinants in the plaque assay, as they formed bigger plaques than the P0 viruses (see **Fig 3.8**).

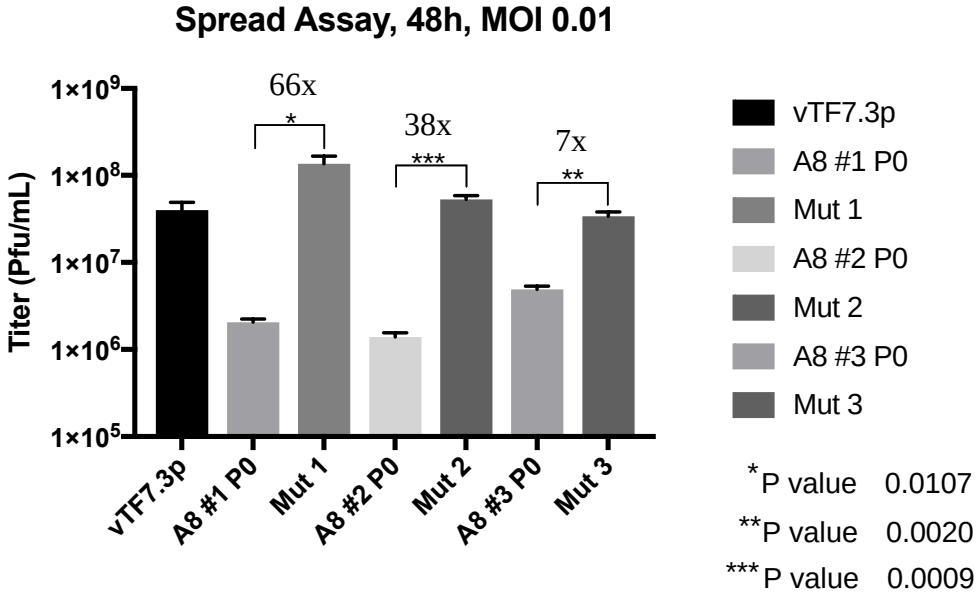

Figure A1. Viral spread of MYXA8 P0 and P10 viruses. BS-C-1 cells were infected in triplicate at an MOI of 0.01. Samples were harvested at 48 hpi and then titrated in duplicate on BS-C-1 cells. Bars represent SEM. This experiment was only performed once.

A2.2 Gene Expression

A.2.2.1 Deletion and P0 Viruses

One of the simplest ways we thought to check how these adaptive mutations worked was by checking protein expression for both intermediate/late viral proteins and the MYXV A8 ortholog. Increased production of these proteins seemed to be the most obvious way these viruses would have a growth advantage over the P0 viruses. First, we checked the expression level of D13 for the initial deletion and P0 viruses used. BS-C-1 cells were infected at an MOI of 5 and samples were harvested at the indicated times. All the viruses gave the expected results, as shown in **Fig A2, top.** The deletion viruses did not express D13, WT expressed normal levels of the protein, and the P0 viruses expressed intermediate levels of the protein, regardless of the timepoint. We also wanted to check the level of expression of the A8 ortholog, via the V5 tag. As expected, the only lane which had a band was that of the Myx A8 P0 (**Fig. A2, middle**).

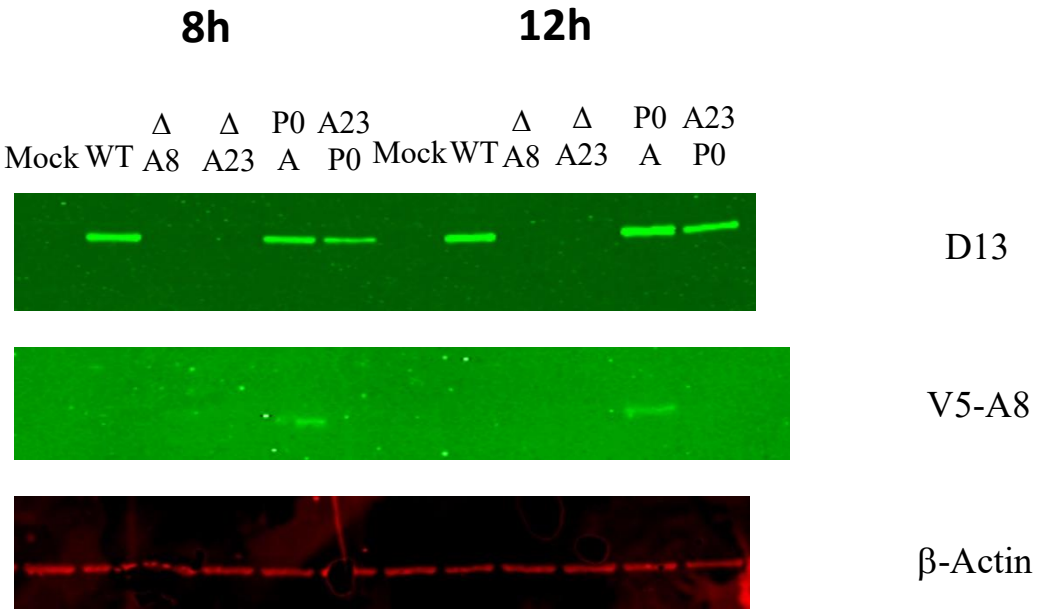

Figure A2. Protein expression of deletion and P0 viruses. Expression of the viral D13 protein (top). BS-C-1 cells were infected at an MOI of 5. Cells were harvested at the indicated times. Antibody against D13 was used to visualize the protein. Expression of the MYXA8 protein (middle). BS-C-1 cells were infected at an MOI of 5. Cells were harvested at the indicated times. Antibody against the V5 tag was used to visualize the protein. Actin was used as a loading control (bottom). This experiment was only performed once.

A2.2.2 P10-isolated Viruses

We wanted to check the protein expression levels for the P10 isolates. First, we checked the expression level of D13. BS-C-1 cells were infected at an MOI of 5 and samples were harvested at the indicated times for both blots. The A8 deletion virus again did not express D13, as expected. D13 expression for WT, P0, and the P10 isolates were similar at both timepoints (**Fig A3, top**). We then wanted to check the MYXA8 expression level. Being an ITF, A8 would likely play a more direct role in the increased expression of intermediate genes, directly, and late genes, indirectly. As predicted, due to the improved Kozak sequence, we saw increased expression of MYXA8 for the P10B clone (**Fig A3, middle**). This supported our hypothesis regarding the MYXA8 mutation in P10B (see **Fig. 3.7A**) resulting in increased expression of the A8 ortholog. At the 8 h time point, we also saw an increase in MYXA8 expression in P10C, which does not have a MYXA8 mutation. This experiment needs to be repeated since we did not see an increase at 12 h. We were unable to detect the V5-tagged ortholog in P10A, as its MYXA8 mutation rendered the epitope tag useless (see **Fig. 3.7A**).

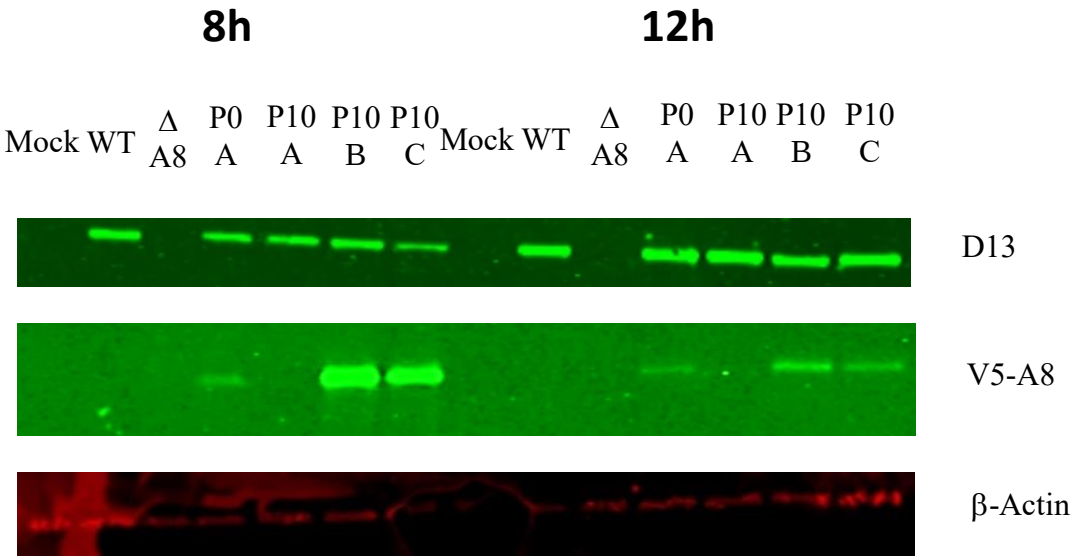

Figure A3. Protein expression of MYXA8 P0 and P10 viruses. Expression of the viral D13 protein (top). BS-C-1 cells were infected at an MOI of 5. Cells were harvested at the indicated times. Antibody against D13 was used to visualize the protein. Expression of the MYXA8 protein (middle). BS-C-1 cells were infected at an MOI of 5. Cells were harvested at the indicated times. Antibody against the V5 tag was used to visualize the protein. Actin was used as a loading control (bottom). This experiment was only performed once.

A2.2.3 Recombinant Viruses

We also wanted to see if there were any changes at the transcriptional level by using digital droplet PCR (ddPCR). The recombinant viruses had similar levels of I3 transcripts at both timepoints and were about the same or lower than the levels of I3 transcripts for the P0 and WT viruses (**Fig A4, top**). They had the same or lower levels of D13 transcripts compared to the P0 virus, but those levels were not as high as wild-type (**Fig. A4, bottom**).

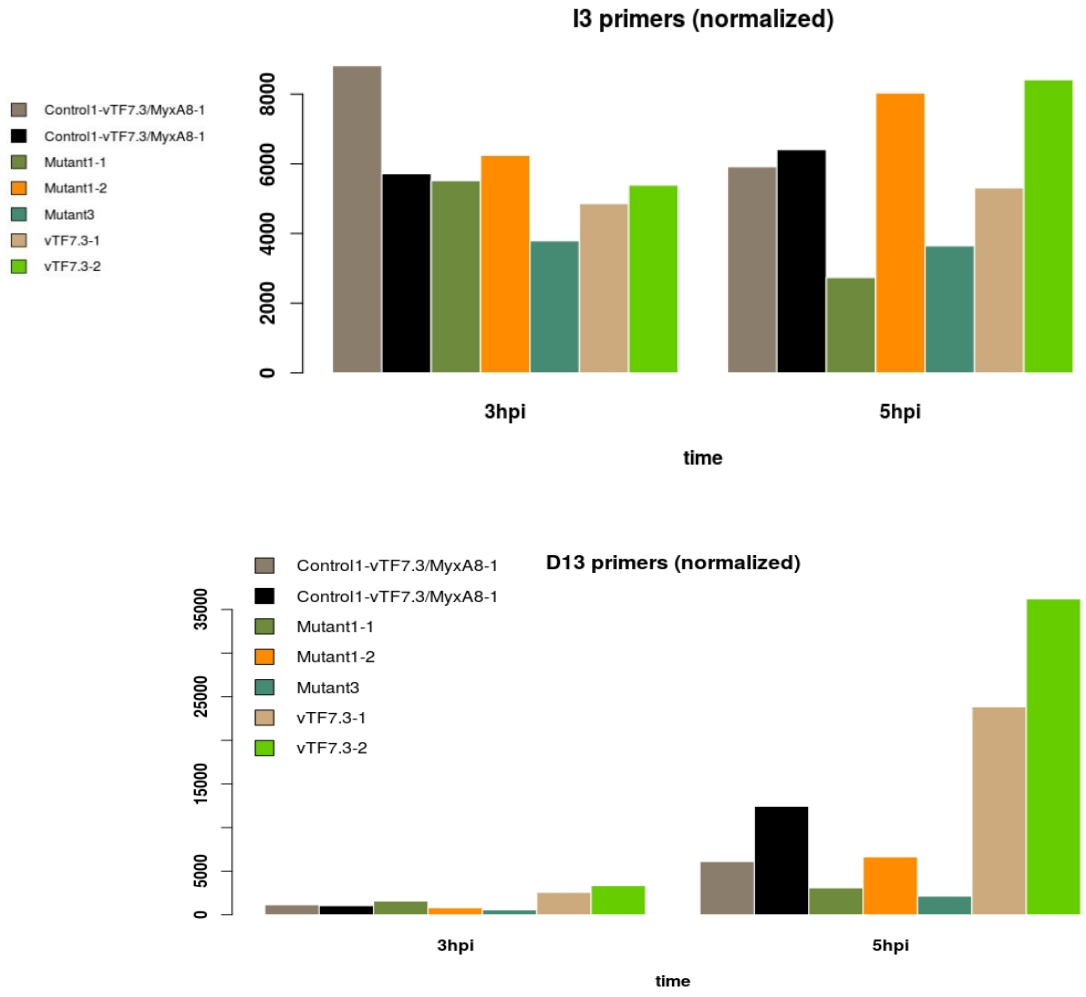

Figure A4. Copy number of I3 and D13 transcripts. BS-C-1 cells were infected at a high MOI and then harvested at the indicated times. RNA was purified and reverse transcribed. PCR was performed using primers specific to I3 (top) and D13 (bottom). Results were normalized to rRNA transcript levels. Each sample was assayed in duplicate, with the exception of Mutant3. This experiment was only performed once.

A2.3 IBT Resistance

A2.3.1 Recombinant Viruses

We then wanted to test if the RNAP mutations specifically granted the virus the ability to counter the effects of IBT. We quantified how well viruses could grow in the presence of the drug. Both the P0 and wt viruses were equally sensitive to the drug, with wt growing only slightly better (**Fig. A5**). Mutant1 also grew poorly in the presence of the drug. Mutant3 and B mut (an additional mutation found in RPO 147 (G1145E)) from a new and independent blind passaging experiment, see **Table A1**) performed much better in the presence of the drug. None of these viruses, however, grew as well as the virus with a known IBT-resistant mutation in RPO132, known as IBT^{r90} (A24R Y462H), first described by Condit et al. 1991 (120).

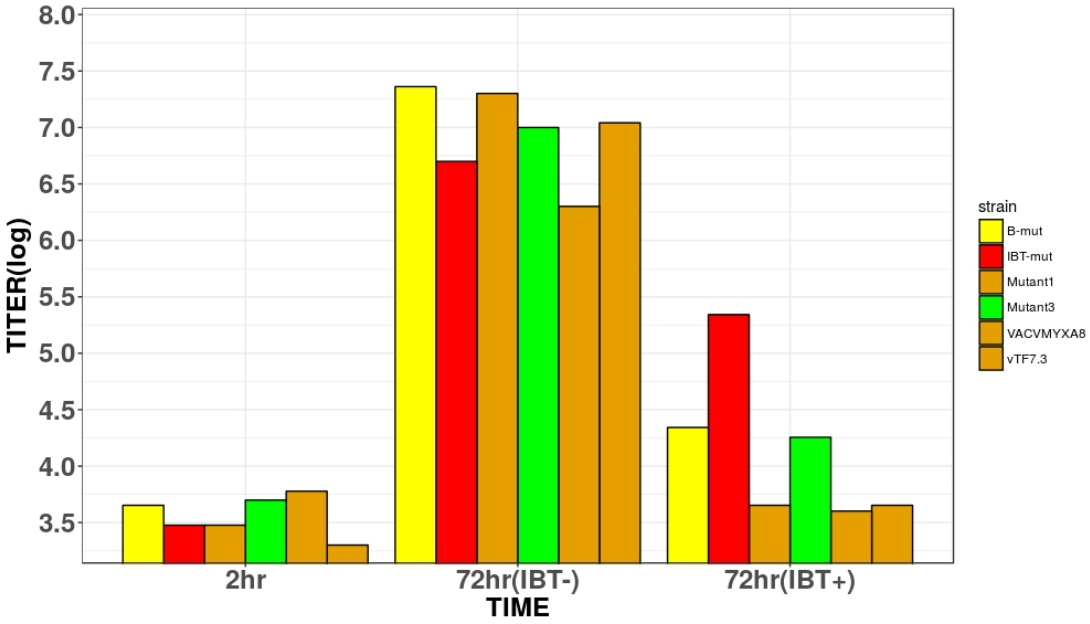

Figure A5. IBT resistance of RNAP mutations. BS-C-1 cells were infected at an MOI of 0.01 either in the presence or absence of 60 μM of IBT. Samples were harvested at the indicated times and then titrated on BS-C-1 cells. This experiment was only performed once.

A2.4 Identification of Additional Mutations

A2.4.1 Blind Passaging

The MYXA8 P0 virus was then subjected to three new independent blind passaging experiments. This was done to potentially identify new adaptive mutations. Frequencies of putative adaptive mutations at various passages in their respective populations were determined by Sanger Sequencing and WGS (data not shown). The percent frequencies of potential adaptive mutations in P10 for all three experiments, as determined by WGS, are shown in **Table A1**.

Table A1. Percent Frequencies of Additional Putative Adaptive Mutations

	Insertion T in F10	Deletion T G6 (frameshift)	Deletion T A14.5	Deletion T A26
p10A	8%	12.6%	74%	98%
	G1145E RPO147 (B Mut)	Deletion 11nt G6	Deletion T G6	Deletion T A26
p10B	18%	61%	15.8%	98%
	Deletion T G6	Deletion T A26	Kozak -3 C-A	
p10C	91%	96%	16%	

A2.5 Materials and Methods

A2.5.1 Spread Assay

BS-C-1 cells were infected at an MOI of 0.01. Samples were harvested at 48 hpi. Cells were lysed by 3 freeze-thaw cycles and sonicated on ice. Samples were then titrated on BS-C-1 cells.

A2.5.2 Western blotting

BS-C-1 cells were infected at an MOI of 5. Samples were harvested at the indicated times. Anti-D13 and V5 tag antibodies were used to detect expression of viral proteins. Fluorescently-labelled secondary antibodies were used (Li-Cor, 1:10,000). Bands were visualized using the Li-Cor Odyssey CLx.

A2.5.3 ddPCR

BS-C-1 cells were infected at an MOI of 5, in duplicate. Samples were harvested at the indicated times. The RNA was purified and then reverse transcribed to obtain cDNA. These samples were then used as templates for PCR using the gene-specific primers indicated. Samples were then put on the BioRad Automated Droplet Generator and read on the BioRad QX200 Droplet Reader.

A2.5.4 IBT Resistance

BS-C-1 cells were infected in duplicate at an MOI of 0.01, washed after 2h, and then fresh media, with or without 60 μM IBT, was added to each well. Samples were harvested at the indicated times and then titrated on BS-C-1 cells.

A2.5.5 Statistical Analysis

Statistical analysis was carried out using Prism GraphPad (Version 7.0c). Unpaired two-tailed T tests were performed on the indicated data. P-values <0.05 were deemed statistically significant.

Appendix B: Previously Published Works

UNIVERSITY OF MARYLAND
BIOLOGICAL SCIENCES GRADUATE PROGRAM

2101 Bioscience Research Building
College Park, Maryland 20742-4415
301.405.6905/6991 TEL, 301.314.9921 FAX

Steve Fetter, Ph.D.
Associate Dean of the Graduate School
University of Maryland

Dear Dr. Fetter,

This letter is written to signify that the dissertation committee, committee chair, and the graduate director have all approved the use of previously published co-authored work in the final dissertation of Carey Stuart, Biological Sciences Graduate Program, 109442794. In accordance with the Graduate School's policy the dissertation committee has determined that they made substantial contributions to the included work.

The citations for the published works are:
Stuart, C.A., Zhivkoplias, E.V., Senkevich, T.G., Wyatt, L.S., and Moss, B. 2018. RNA polymerase mutations selected during experimental evolution enhance replication of a hybrid vaccinia virus with an intermediate transcription factor subunit replaced by the myxoma virus ortholog. J Virol 92(20): e01089-18.

Per Graduate School policy the dissertation forward will identify the scope and nature of the student's contributions to the jointly authored work included in the dissertation and a copy of this letter will be submitted with the dissertation.

Sincerely,

Dr. Jeffrey DeStefano, Dissertation Committee Chair, Professor, Molecular and Cellular Biology

Dr. Zakiya Whatley, Program Manager, Biological Sciences Graduate Program

Carey Stuart, Graduate Student, Biological Sciences

Bibliography

1. Babkin, I.V. and Babkina, I.N., *The Origin of the Variola Virus.* Viruses, 2015. **7**: p. 100-1112.

2. Colson, P., et al., *"Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses.* Arch Virol, 2013. **158**(12): p. 2517-2521.

3. Yang, Z., et al., *Expression Profiling of the Intermediate and Late Stages of Poxvirus Replication.* J Virol, 2011. **85**(19): p. 9899-9908.

4. Mirzakhanyan, Y. and Gerson, P.D., *Multisubunit DNA-Dependent RNA Polymerases from Vaccinia Virus and Other Nucleocytoplasmic Large-DNA Viruses: Impressions from the Age of Structure.* Microbiol Mol Biol Rev, 2017. **81**(3): p. e00010-17.

5. Walsh, D., *Poxviruses: Slipping and sliding through transcription and translation.* PLoS Pathog, 2017. **13**(11): p. e1006634.

6. Weisberg, A.S., et al., *Enigmatic Origin of the poxvirus membrane from the endoplasmic reticulum shown by 3D imaging of vaccinia virus assembly mutants.* Proc Natl Acad Sci USA, 2017. **114**(51): p. E11001-E11009.

7. Moss, B., *Poxviridae: The Viruses and Their Replication*, in *Fields Virology*, P.M. Howley, Editor. 2007, Lippincott Williams & Wilkins. p. 2849-2883.

8. Shchelkunov, S.N., *Emergence and reemergence of smallpox: The need for development of a new generation smallpox vaccine.* Vaccine, 2011: p. D49-D53.

9. Theves, C., Biagini, P., and Crubezy, E., *The rediscovery of smallpox.* Clin Microbiol Infect, 2014. **20**: p. 210-218.

10. Flint, S.J., et al., *Principles of Virology*. 4th ed. Pathogenesis and Control. Vol. II. 2015, Washington, D.C.: ASM Press.

11. Sanchez-Sampedro, L., et al., *The Evolution of Poxvirus Vaccines.* Viruses, 2015. **7**(4): p. 1726-1803.

12. Medaglia, M.L.G., et al., *Genomic Analysis, Phenotype, and Virulence of the Historical Brazilian Smallpox Vaccine Strain IOC: Implications for the Origins and Evolutionary Relationships of Vaccinia Virus.* J Virol, 2015. **89**(23): p. 11909-11925.

13. Esparza, J., et al., *Equination (inoculation of horsepox): An early alternative to vaccination (inoculation of cowpox) and the potential role of horsepox virus in the origin of the smallopx vaccine.* Vaccine, 2017. **35**: p. 7222-7230.

14. Schrick, L., et al., *An Early American Smallpox Vaccine Based on Horsepox.* The New England Journal of Medicine, 2017. **377**(15).

15. Li, Y., et al., *On the orgin of smallpox: Correlating variola phylogenics with historical smallpox records.* Proc Natl Acad Sci USA, 2007. **104**(40): p. 15787-15792.

16. Damaso, C.R., *Revisiting Jenner's mysteries, the role of the Beaugency lymph in the evolutionary path of ancient smallpox vaccines.* Lancet Infect Dis, 2018. **18**: p. e55-63.

17. Brown, D. *Rinderpest, or 'cattle plaque,' becomes only the second disease to be eradicated*. The Washington Post, 2011.

18. McNeil, J., D.G. *Rinderpest, Scourge of Cattle, Is Vanquished*. The New York Times, 2011.

19. Roeder, P., Mariner, J., and Kock, R., *Rinderpest: the veterinary perspective on eradication.* Phil Trans R Soc B, 2013. **368**(1623): p. 20120139.

20. *Bulgaria reports first outbreak of ovine rinderpest in EU*. Reuters, 2018.

21. Berezow, A. *Ovine Rinderpest Outbreak in Europe: Wait, Wasn't That Eradicated?* American Council on Science and Health, 2018.

22. Flint, S.J., et al., *Principles of Virology*. 3rd ed. Pathogenesis and Control. Vol. II. 2009, Washington, D.C.: ASM Press.

23. Tulman, E.R., et al., *Genome of Horsepox Virus.* J Virol, 2006. **80**(18): p. 9244-9258.

24. Esparza, J., *Has horsepox become extinct?* Veterinary Record, 2013. **173**: p. 272-273.

25. Flint, S.J., et al., *Principles of Virology*. 3rd ed. Molecular Biology. Vol. I. 2009, Washington, D.C.: ASM Press.

26. Trovato, M. and Berardinia, P.D., *Novel antigen delivery systems.* World J Virol, 2015. **4**(3): p. 156-168.

27. Moss, B., *Reflections on the Early Development of Poxvirus Vectors.* Vaccine, 2013. **31**(39): p. 4220-4222.

28. Flint, S.J., et al., *Principles of Virology*. 4th ed. Molecular Biology. Vol. I. 2015, Washington, D.C.: ASM Press.

29. Kaufman, H.L., Kohlhapp, F.J., and Zloza, A., *Oncolytic viruses: a new class of immunotherapy drugs.* Nature Reviews, 2015. **14**: p. 642-662.

30. Villa, N.Y., et al., *Myxoma virus suppresses proliferation of activated T lymphocytes yet permits oncolytic viurs transfer to cancer cells.* Blood, 2015. **125**(24): p. 3778-3788.

31. Conrad, S.J., et al., *Oncolytic tanapoxvirus expressing FliC causes regression of human colorectal cancer xenografts in nude mice.* Journal of Experimental and Clinical Cancer Research, 2015. **34**(19).

32. Tulman, E.R., et al., *The Genome of Canarypox Virus.* J Virol, 2004. **78**(1): p. 353-366.

33. Parviainen, S., et al., *GMCSF-armed vaccinia virus induces an antitumor immune response.* Int. J. Cancer, 2015. **136**: p. 1065-1072.

34. Nelson, C.A., et al., *Structural Conservation and Functional Diversity of the Poxvirus Immune Evasion (PIE) Domain Superfamily.* Viruses, 2015. **7**: p. 4878-4898.

35. Bidgood, S.R. and Mercer, J., *Cloak and Dagger: Alternative Immune Evasion and Modulation Strategies of Poxviruses.* Viruses, 2015. **7**: p. 4800-4825.

36. Prideaux, C.T., Kumar, S., and Boyle, D.B., *Comparative analysis of vaccinia virus promoter activity in fowlpox and vaccinia virus recombinants.* Virus Res, 1990. **16**: p. 43-58.

37. Tripathy, D.N. and Wittek, R., *Regulation of foreign gene in fowlpox virus by a vaccinia virus promoter.* Avian Dis, 1990. **34**(1): p. 218-220.

38. Kumar, S. and Boyle, D.B., *Activity of a fowlpox virus late gene promoter in vaccinia and fowlpox virus recombinants.* Arch Virol, 1990. **112**(3-4): p. 139-48.

39. Mutsafi, Y., et al., *Vaccinia-like cytoplasmic replication of the giant Mimivirus.* Proc Natl Acad Sci USA, 2010. **107**(13): p. 5978-5982.

40. Moss, B., *Poxviridae: The Viruses and Their Replication*, in *Fields Virology*, P.M. Howley and D.M. Knipe, Editors. 2013, Lippincott Williams & Wilkins. p. 2129-2159.

41. Afonso, C.L., et al., *Genome of Crocodilepox Virus.* J Virol, 2006. **80**(10): p. 4978-4991.

42. Hendrickson, R.C., et al., *Orthopoxvirus Genome Evolution: The Role of Gene Loss.* Viruses, 2010. **2**: p. 1933-1967.

43. Hughes, A.L., Irausquin, S., and Friedman, R., *The Evolutionary Biology of Poxviruses.* Infect Genet Evol, 2010. **10**(1).

44. Babkin, I.V. and Babkina, I.N., *A retrospective study of the orthopoxvirus molecular evolution.* Infect Genet Evol, 2012. **12**: p. 1597-1604.

45. Smithson, C., et al., *Prediction of Steps in the Evolution of Variola Virus Host Range.* PLoS ONE, 2014. **9**(3): p. e91520.

46. Shchelkunov, S.N., *How long ago did smallpox virus emerge?* Arch Virol, 2009. **154**: p. 1865-1871.

47. Laliberte, J.P. and Moss, B., *Lipid Membranes in Poxvirus Replication.* Viruses, 2010. **2**: p. 972-986.

48. Moss, B., *Poxvirus Cell Entry: How Many Proteins Does it Take?* Viruses, 2012. **4**: p. 688-707.

49. Laliberte, J.P., Weisberg, A.S., and Moss, B., *The Membrane Fusion Step of Vaccinia Virus Entry is Cooperatively Mediated by Viral Proteins and Host Cell Components.* PLoS Pathog, 2011. **7**(12).

50. Moss, B., *Membrane fusion during poxvirus entry.* Semin Cell Dev Biol, 2016. **60**: p. 89-96.

51. Grossegesse, M., et al., *A Next-Generation Sequencing Approach Uncovers Viral Transcripts Incorporated in Poxvirus Virions.* Viruses, 2017. **9**(296).

52. Maruri-Avidal, L., Weisberg, A.S., and Moss, B., *Vaccinia Virus L2 Protein Associates with the Endoplasmic Reticulum near the Growing Edge of Crescent Precursors of Immature Virions and Stabilizes a Subset of Viral Membrane Proteins.* J Virol, 2011. **85**(23): p. 12431-12441.

53. Maruri-Avidal, L., et al., *Analysis of Viral Membranes Formed in Cells Infected by a Vaccinia Virus L2-Deletion Mutant Suggests Their Origin from the Endoplasmic Reticulum.* J Virol, 2013. **87**(3): p. 1861-1871.

54. Maruri-Avidal, L., Weisberg, A.S., and Moss, B., *Association of the Vaccinia Virus A11 Protein with the Endoplasmic Reticulum and Crescent Precursors of Immature Virions.* J Virol, 2013. **87**(18): p. 10195-10206.

55. Maruri-Avidal, L., Weisberg, A.S., and Moss, B., *Direct Formation of Vaccinia Virus Membranes from the Endoplasmic Reticulum in the Absence of the Newly Characterized L2-Interacting Protein A30.5.* J Virol, 2013. **87**(22): p. 12313-12326.

56. Lefkowitz, E.J., Wang, C., and Upton, C., *Poxviruses: past, present and future.* Virus Research, 2006: p. 105-118.

57. Upton, C., et al., *Poxvirus Orthologous Clusters: toward Defining the Minimum Essential Poxvirus Genome.* J Virol, 2003. **77**(13): p. 7590-7600.

58. McLysaght, A., Baldi, P.F., and Gaut, B.S., *Extensive gene gain associated with adaptive evolution of poxviruses.* Proc Natl Acad Sci USA, 2003. **100**(26): p. 15655-15660.

59. Baroudy, B.M., Venkatesan, S., and Moss, B., *Incompletely Base-Paired Flip-Flop Terminal Loops Link the Two DNA Strands of the Vaccinia Virus Genome into One Uninterrupted Polynucleotide Chain.* Cell, 1982. **28**: p. 315-324.

60. Baroudy, B.M., Venkatesan, S., and Moss, B., *Structure and Replication of Vaccinia Virus Telomeres.* Cold Spring Harbor Symposia on Quantitative Biology, 1983. **XLVII**: p. 723-729.

61. Merchlinksy, M., Garon, C.F., and Moss, B., *Molecular Cloning and Sequence of the Concatemer Junction from Vaccinia Virus Replicative DNA.* J Mol Biol, 1988. **199**: p. 399-413.

62. Du, S. and Traktman, P., *Vaccinia virus DNA replication: Two hundred base pairs of telomeric sequence confer optimal replication efficiency on minichromosome templates.* Proc Natl Acad Sci USA, 1996. **93**: p. 9693-9698.

63. DeFilippes, F.M., *Restriction Enzyme Mapping of Vaccinia Virus DNA.* J Virol, 1982. **43**(1): p. 136-149.

64. Muller, H.K., et al., *Comparison of Five Poxvirus Genomes by Analysis with Restriction Endonucleases HindIII, BamI and EcoRI.* J Gen Virol, 1977. **38**(135-147).

65. Bengali, Z., et al., *Drosophila S2 cells are non-permissive for vaccinia virus DNA replication following entry via low pH-dependent endocytosis and early transcription.* PLoS ONE, 2011. **6**(2): p. e17248.

66. Hyun, S.I., Weisberg, A.S., and Moss, B., *Deletion of the Vaccinia Virus I2 Protein Interrupts Virion Morphogenesis, Leading to Retention of the Scaffold Protein and Mislocalization of Membrane-Associated Entry Proteins.* J Virol, 2017. **91**(15): p. e00558-17.

67. Satheshkumar, P.S. and Moss, B., *Sequence-Divergent Chordopoxvirus Homologs of the O3 Protein Maintain Functional Interactions with*

Components of the Vaccinia Virus Entry-Fusion Complex. J Virol, 2011. **86**(3): p. 1696-1705.

68. Laliberte, J.P. and Moss, B., *Appraising the apoptotic mimicry model and the role of phospholipids for poxvirus entry.* Proc Natl Acad Sci USA, 2009. **106**(41): p. 17517-17521.

69. Brennan, G., et al., *Experimental Evolution Identifies Vaccinia Virus Mutations in A24R and A35R That Antagonize the Protein Kinase R Pathway and Accompany Collapse of an Extragenic Gene Amplification.* J Virol, 2015. **89**(19): p. 9986-9997.

70. Mendez-Rios, J.D., et al., *Molluscum Contagiosum Virus Transcriptome in Abortively Infected Cultured Cells and a Human Skin Lesion.* J Virol, 2016. **90**(9): p. 4469-4480.

71. Laliberte, J.P. and Moss, B., *A Novel Mode of Poxvirus Superinfection Exclusion That Prevents Fusion of the Lipid Bilayers of Viral and Cellular Membranes.* J Virol, 2014. **88**(17): p. 9751-9768.

72. Senkevich, T., et al., *Identification of Vaccinia Virus Replisome and Transcriptome Proteins by Isolation of Proteins on Nascent DNA Coupled with Mass Spectrometry*. J Virol, 2017. **91**(19): p. e01015-17.

73. Liu, B., et al., *Identification of Poxvirus Genome Uncoating and DNA Replication Factors with Mutually Redundant Roles.* J Virol, 2018.

74. Senkevich, T., et al., *Mapping vaccinia virus DNA replication origins at nucleotide level by deep sequencing.* Proc Natl Acad Sci USA, 2015. **112**(35): p. 10908-10913.

75. De Silva, F.S., et al., *Poxvirus DNA Primase.* Proc Natl Acad Sci USA, 2007. **104**(47): p. 18724-18729.

76. Paran, N., et al., *Cellular DNA Ligase I is Recruited to Cytoplasmic Vaccinia Virus Factories and Masks the Role of the Vaccinia Ligase in Viral DNA Replication.* Cell Host Microbe, 2009. **6**(6): p. 563-569.

77. Katsafanas, G.C., et al., *Inhibition of Vaccinia Virus Replication by N-(Phosphonoacetyl)-L-aspartate: Differential Effects on Viral Gene Expression Result from a Reduced Pyrimidine Nucleotide Pool.* Virology, 1997. **236**: p. 177-187.

78. Moss, B., *Poxvirus DNA Replication.* Cold Spring Harb Perspect Biol, 2013. **5**: p. a010199.

79. Yutin, N., et al., *Chordopoxvirus protein F12 implicated in enveloped virion morphogenesis is an inactivated DNA polymerase.* Biology Direct, 2014. **9**(22).

80. Merchlinksy, M. and Moss, B., *Resolution of Vaccinia Virus DNA Concatemer Junctions Requires Late-Gene Expression.* J Virol, 1989. **63**(4): p. 1595-1603.

81. Garcia, A.D., et al., *Bacterial-type DNA Holliday junction resolvases in eukaryotic viruses.* Proc Natl Acad Sci USA, 2000. **97**(16): p. 8926-8931.

82. Garcia, A.D. and Moss, B., *Repression of Vaccinia Virus Holliday Junction Resolvase Inhibits Processing of Viral DNA into Unit-Length Genomes.* J Virol, 2001. **75**(14): p. 6460-6471.

83. Fonseca, F.D. and Moss, B., *Poxvirus DNA topoisomerase knockout mutant exhibits decreased infectivity associated with reduced early transcription.* Proc Natl Acad Sci USA, 2003. **100**(20): p. 11291-11296.

84. Senkevich, T., Koonin, E., and Moss, B., *Predicted poxvirus FEN1-like nuclease required for homologous recombination, double-strand break repair and full-size genome formation.* Proc Natl Acad Sci USA, 2009. **106**(42): p. 17921-17926.

85. Cassetti, M.C., et al., *DNA Packaging Mutant: Repression of the Vaccinia Virus A32 Gene Results in Noninfectious, DNA-Deficient, Spherical, Enveloped Particles.* J Virol, 1998. **72**(7): p. 5769-5780.

86. Panda, D., et al., *Triad of human cellular proteins, IRF2, FAM111A, and RFC3, restrict replication of orthopoxvirus SPI-1 host-range mutants.* Proc Natl Acad Sci USA, 2017. **114**(14): p. 3720-3725.

87. Moss, B., *Poxvirus Membrane Biogenesis.* Virology, 2015. **0**: p. 619-626.

88. Moss, B., *Origin of the poxviral membrane: A 50-year-old riddle.* PLoS Pathog, 2018. **14**(6): p. e1007002.

89. Hyun, S.I., Maruri-Avidal, L., and Moss, B., *Topology of Endoplasmic Reticulum-Associated Cellular and Viral Proteins Determined with Split-GFP*. Traffic, 2015. **16**(7): p. 787-795.

90. Maruri-Avidal, L., et al., *Participation of Vaccinia Virus L2 Protein in the Formation of Crescent Membranes and Immature Virions.* J Virol, 2011. **85**(6): p. 2504-2511.

91. Erlandson, K.J., et al., *Poxviruses Encode a Reticulon-Like Protein that Promotes Membrane Curvature.* Cell Rep, 2016. **14**(9): p. 2084-2091.

92. McFadden, B.D.H., et al., *Vaccinia virions deficient in transcription enzymes lack a nucleocapsid.* Virology, 2012. **434**: p. 50-58.

93. Sivan, G., et al., *Human genome-wide RNAi screen reveals a role for nuclear pore proteins in poxvirus morphogenesis.* Proc Natl Acad Sci USA, 2013. **110**(9): p. 3519-3524.

94. Kastenmayer, R.J., et al., *Elimination of A-type inclusion formation enhances cowpox virus replication in mice: implications for orthopoxvirus evolution.* Virology, 2014. **452-453**: p. 59-66.

95. Sivan, G., et al., *Retrograde Transport from Early Endosomes to the trans-Golgi Network Enables Membrane Wrapping and Egress of Vaccinia Virus Virions.* J Virol, 2016. **90**(19): p. 8891-8905.

96. Byrd, C.M. and Hruby, D.E., *Screening for Vaccinia Virus Egress Inhibitors: Separation of IMV, IEV, and EEV*, in *Methods in Molecular Biology*, S.N. Isaacs, Editor. 2012, Springer Sciene+Business Media, LLC. p. 113-120.

97. Yang, Z., et al., *Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing.* Proc Natl Acad Sci USA, 2010. **107**: p. 11513-11518.

98. Yang, Z., et al., *Cascade regulation of vaccinia virus expression is modulated by multistage promoters.* Virology, 2013. **447**(1-2): p. 213-220.

99. Parrish, S., et al., *The African Swine Fever Virus g5R Protein Possesses mRNA Decapping Activity.* Virology, 2009. **393**(1): p. 177-182.

100. Liu, S.W., et al., *The D10 Decapping Enzyme of Vaccinia Virus Contributes to Decay of Cellular and Viral mRNAs and to Virulence in Mice.* J Virol, 2014. **88**(1): p. 202-211.

101. Liu, S.W., et al., *Poxvirus Decapping Enzymes Enhance Virulence by Preventing the Accumulation of dsRNA and the Induction of Innate Antiviral Repsonses.* Cell Host and Microbe, 2015. **17**(3): p. 320-331.

102. Liu, R. and Moss, B., *Opposing Roles of Double-Stranded RNA Effector Pathways and Viral Defense Proteins Revealed with CRISPR-Cas9 Knockout Cell Lines and Vaccinia Virus Mutants.* J Virol, 2016. **90**(17): p. 7864-7879.

103. Dai, A., et al., *Ribosome profiling reveals translational upregulation of cellular oxidative phosphorylation mRNAs during vaccinia virus-induced host shutoff.* J Virol, 2017. **91**(5): p. e01858-16.

104. Cao, S., Dhungel, P., and Yang, Z., *Going against the tide: selective cellular protein synthesis during virall induced host shutoff.* J Virol, 2017. **91**(17): p. e00071-17.

105. Katsafanas, G.C. and Moss, B., *Colocalization of Transcription and Translation within Cytoplasmic Poxvirus Factories Coordinates Viral Expression and Subjugates Host Functions.* Cell Host and Microbe, 2007. **2**: p. 221-228.

106. Thompson, M.R., et al., *Pattern Recognition Receptors and the Innate Immune Response to Viral Infection.* Viruses, 2011. **3**: p. 920-940.

107. Yang, Z., et al., *Genome-wide analysis of the 5' and 3' ends of vaccinia early mRNAs delineates regulatory sequences of annotated and anomalous transcripts.* J Virol, 2011. **85**(12): p. 5897-5909.

108. Broyles, S.S., *Vaccinia virus transcription.* J Gen Virol, 2003. **84**: p. 2293-2303.

109. Broyles, S.S., Li, J., and Moss, B., *Promoter DNA contacts made by the vaccinia early transcription factor.* J Biol Chem, 1991. **266**(23): p. 15539-15544.

110. Yang, Z. and Moss, B., *Interaction of the Vaccinia Virus RNA Polymerase-Associated 94-Kilodalton Protein with the Early Transcription Factor.* J Virol, 2009. **83**(23): p. 12018-12026.

111. Tate, J. and Gollnick, P., *The role of vaccinia termination factor and cis-acting elements in vaccinia virus early gene transcription termination.* Virology, 2015. **485**: p. 179-188.

112. Daffis, S., et al., *2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members.* Nature, 2010. **468**(7322): p. 452-456.

113. Hirschmann, P., Vos, J.C., and Stunnenberg, H.G., *Mutation Analysis of a Vaccinia Virus Intermediate Promoter In Vivo and In Vitro.* J Virol, 1990. **64**(12): p. 6063-6069.

114. Vos, J.C., Sasker, M., and Stunnenberg, H.G., *Promoter Melting by a Stage-Specific Vaccinia Virus Transcription Factor Is Independent of the Presence of RNA Polymerase.* Cell, 1991. **65**: p. 105-113.

115. Keck, J.G., Baldick, J., C.J., and Moss, B., *Role of DNA Replication in Vaccinia Virus Gene Expression: A Naked Template Is Required for Transcription of Three Late Trans-Activator Genes.* Cell, 1990. **61**: p. 801-809.

116. Vos, J.C., Sasker, M., and Stunnenberg, H.G., *Vaccinia virus capping enzyme is a transcription initiation factor.* The EMBO Journal, 1991. **10**(9): p. 2553-2558.

117. Sanz, P. and Moss, B., *Identification of a transcription factor, encoded by two vaccinia virus early genes, that regulates the intermediate stage of viral gene expression.* Proc Natl Acad Sci USA, 1999. **96**(6): p. 2692-2697.

118. Warren, R.D., Cotter, C.A., and Moss, B., *Reverse genetic analysis of poxvirus intermediate transcription factors.* J Virol, 2012. **86**(17): p. 6470-6479.

119. Katsafanas, G.C. and Moss, B., *Vaccinia Virus Intermediate Stage Trancription Is Complemented by Ras-GTPase-activating Protein SH3 Domain-binding Protein (G3BP) and Cytoplasmic Activation/Proliferation-associated Protein (p137) Individually or as a Heterodimer.* The Journal of Biological Chemistry, 2004. **279**(50): p. 52210-52217.

120. Condit, R.C., et al., *A Vaccinia Virus Istatin-B-Thiosemicarbazone Resistance Mutation Maps in the Viral Gene Encoding the 132-kDa Subunit of RNA Polymerase.* Virology, 1991. **185**(2): p. 857-861.

121. Prins, C., Cresawn, S.G., and Condit, R.C., *An Isatin-B-thiosemicarbazone-resistant Vaccinia Virus Containing a Mutation in the Second Largest Subunit of the Viral RNA Polymerase is Defective in Transcription Elongation.* The Journal of Biological Chemistry, 2004. **279**(43): p. 44858-44871.

122. Cresawn, S.G., et al., *Mapping and phenotypic analysis of spontaneous isatin-B-thiosemicarbazone resistant mutants of vaccinia virus.* Virology, 2007. **363**(2): p. 319-332.

123. Black, E.P., Moussatche, N., and Condit, R.C., *Characterization of the Interactions among Vaccinia Virus Transcription Factors G2R, A18R, and H5R.* Virology, 1998. **245**: p. 313-322.

124. Dhungel, P., Cao, S., and Yang, Z., *The 5'-poly(A) leader of poxvirus mRNA confers a translation advantage that can be achieved in cells with impaired cap-dependent tranlsation.* PLoS Pathog, 2017. **13**(8): p. e1006602.

125. Yang, Z., et al., *Pervasive Initiation and 3'-End Formation of Poxvirus Postreplicative RNAs.* The Journal of Biological Chemistry, 2012. **287**(37): p. 31050-31060.

126. Fuerst, T.R., et al., *Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase.* Proc Natl Acad Sci USA, 1986. **83**(21): p. 8122-8126.

127. Moss, B. and Earl, P.L., *Overview of the vaccinia virus expression system*, in *Current Protocols in Molecular Biology*. 2002. p. 5.11:5.11.1-5.11.5; 16.15:16.15.1-16.15.5.

128. Dellis, S., et al., *Protein interactions among the vaccinia virus late transcription factors.* Virology, 2004. **329**: p. 328-336.

129. Yang, Z., et al., *Deciphering poxvirus gene expression by RNA sequencing and ribosome profiling.* J Virol, 2015. **89**(13): p. 6874-6886.

130. Yang, Z. and Moss, B., *Decoding poxvirus genome.* Oncotarget, 2015. **6**(30): p. 28513-28514.

131. Filee, J., *Lateral gene transfer, lineage-specific gene expansion and the evolution of Nucleo Cytoplasmic Large DNA viruses.* Journal of Invertebrate Pathology, 2009. **101**: p. 169-171.

132. Knutson, B.A. and Broyles, S.S., *Expansion of poxvirus RNA polymerase subunits sharing homology with corresponding subunits of RNA polymerase II.* Virus Genes, 2008. **36**: p. 307-311.

133. Koonin, E., Senkevich, T., and Dolja, V.V., *The ancient Virus World and evolution of cells.* Biology Direct, 2006. **1**(29).

134. Koonin, E. and Dolja, V.V., *A virocentric perspective on the evolution of life.* Curr Opin Virol, 2013. **3**(5): p. 546-557.

135. Koonin, E., *The wonder world of microbial viruses.* Expert Rev Anti Infect Ther, 2010. **8**(10): p. 1097-1099.

136. Filee, J., Pouget, N., and Chandler, M., *Phylogenetic evidence for extensive lateral acquisition of cellular genes by Nucleocytoplasmic large DNA viruses.* BMC Evolutionary Biology, 2008. **8**(320).

137. Koonin, E., *Temporal order of evolution of DNA replication systems inferred by comparison of cellular and viral DNA polymerases.* Biology Direct, 2006. **1**(39).

138. Iyer, L.M., et al., *Evolutionary genomics of nucleo-cytoplasmic large DNA viruses.* Virus Res, 2006. **117**: p. 156-184.

139. Koonin, E. and Yutin, N., *Origin and Evolution of Eukaryotic Large Nuceo-Cytoplasmic DNA Viruses.* Intervirology, 2010. **53**: p. 284-292.

140. Koonin, E., Krupovic, M., and Yutin, N., *Evolution of double-stranded DNA viruses of eukaryotes: from bacteriophages to transposons to giant viruses.* Ann NY Acad Sci, 2015. **1341**: p. 10-24.

141. Krupovic, M. and Koonin, E., *Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution.* Nature Reviews Microbiology, 2015. **13**: p. 105-115.

142. Koonin, E., Dolja, V.V., and Krupovic, M., *Origins and evolution of viruses of eukaryotes: The ultimate modularity.* Virology, 2015(479-480): p. 2-25.

143. Filee, J., *Route of NCLDV evolution: the genomic accordion.* Curr Opin Virol, 2013. **3**: p. 595-599.

144. Filee, J., *Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution.* Front. Microbiol, 2015. **6**.

145. Yutin, N., Wolf, Y.I., and Koonin, E., *Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life.* Virology, 2014. **0**: p. 38-52.

146. Elde, N.C., et al., *Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses.* Cell, 2012. **150**(4): p. 831-841.

147. Koonin, E., *On the Origin of Cells and Viruses: Primordial Virus World Scenario.* Ann NY Acad Sci, 2009. **1178**: p. 47-64.

148. Filee, J., et al., *Evolution of DNA Polymerase Families: Evidences for Multiple Gene Exchange Between Cellular and Viral Proteins.* J Mol Evol, 2002. **54**: p. 763-773.

149. Filee, J., Forterre, P., and Laurent, J., *The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies.* Research in Microbiology, 2003. **154**: p. 237-243.

150. Filee, J. and Chandler, M., *Gene Exchange and the Origin of Giant Viruses.* Intervirology, 2010. **53**: p. 354-361.

151. Hughes, A.L. and Friedman, R., *Poxvirus genome evolution by gene gain and loss.* Molecular Phylogenetics and Evolution, 2005. **35**: p. 186-195.

152. Yutin, N. and Koonin, E., *Evolution of DNA ligases of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes: a case of hidden complexity.* Biology Direct, 2009. **4**(51).

153. Yutin, N. and Koonin, E., *Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes.* Virology Journal, 2012. **9**(161).

154. Forterre, P. and Prangishvili, D., *The major role of viruses in cellular evolution: facts and hypotheses.* Curr Opin Virol, 2013. **3**(5): p. 558-565.

155. Filee, J., *Multiple occurrences of giant virus core genes acquired by eukaryotic genomes: The visible part of the iceberg?* Virology, 2014. **466-467**: p. 53-59.

156. La Scola, B., et al., *The virophage as a unique parasite of the giant mimivirus.* Nature Letters, 2008. **455**: p. 100-105.

157. Filee, J., Siguier, P., and Chandler, M., *I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses.* Trends in Genetics, 2006. **23**(1): p. 10-15.

158. Filee, J. and Chandler, M., *Convergent mechanisms of genome evolution of large and giant DNA viruses.* Research in Microbiology, 2008. **159**: p. 325-331.

159. Desnues, C., et al., *Provirophages and transpovirons as the diverse mobilome of giant viruses.* Proc Natl Acad Sci USA, 2012. **109**(44): p. 18078-18083.

160. Odom, M.R., Hendrickson, R.C., and Lefkowitz, E.J., *Poxvirus protein evolution: Family-wide assessment of possible horizontal gene transfer events.* Virus Res, 2009. **144**(1-2): p. 233-249.

161. Oliveira, G.P., et al., *Promoter Motifs in NCLDVs: An Evolutionary Perspective.* Viruses, 2017. **9**(16).

162. Yutin, N., et al., *Eukaryotic large nucleo-cytoplasmic DNA viruses: Clusters of orthologous genes and reconstruction of viral genome evolution.* Virology Journal, 2009. **6**(223).

163. Andrade, A.C.D.S.P., et al., *Filling Knowledge Gaps for Mimivirus Entry, Uncoating, and Morphogenesis.* J Virol, 2017. **91**(22): p. e01335-17.

164. Gjessing, M.C., et al., *Salmon Gill Poxvirus, the Deepest Representative of the Chordopoxvirinae.* J Virol, 2015. **89**(18): p. 9348-9367.

165. Noyce, R.S., Lederman, S., and Evans, D.H., *Construction of an infectious horsepox virus vaccine from chemically synthesized DNA fragments.* PLoS ONE, 2018. **13**(1): p. e0188453.

166. Koblentz, G.D., *The De Novo Synthesis of Horsepox Virus: Implications for Biosecurity and Recommendations for Preveting the Reemergence of Smallpox.* Health Secur, 2017. **15**(6): p. 620-628.

167. DiEuliis, D., Berger, K., and Gronvall, G., *Biosecurity Implications for the Synthesis of Horsepox, an Orthopoxvirus.* Health Secur, 2017. **15**(6): p. 629-637.

168. Kupferschmidt, K., *Critics see only risks, no benefits in horsepox paper.* Science, 2018. **359**(6374): p. 375-376.

169. Koblentz, G.D., *A Critical Analysis of the Scientific and Commercial Rationales for the De Novo Synthesis of Horsepox Virus.* mSphere, 2018. **3**(2): p. e00040-18.

170. DiEuliis, D. and Gronvall, G.K., *A Holistic Assessment of the Risks and Benefits of the Synthesis of Horsepox Virus.* mSphere, 2018. **3**(2): p. e00074-18.

171. Imperiale, M.J., *Re-creation of Horsepox Virus.* mSphere, 2018. **3**(2): p. e00079-18.

172. Snow, J. and Giordano, J., *Public Safety and National Security Implications of the Horsepox Study.* Health Secur, 2018. **16**(2): p. 140-142.

173. DiEuliis, D. and Gronvall, G.K., *Erratum for DiEuliis and Gronvall, "A Holistic Assessment of the Risks and Benefits of the Synthesis of Horsepox Virus.* mSphere, 2018. **3**(2): p. e00160-18.

174. Cone, K.R., et al., *Emergence of a Viral RNA Polymerase Variant during Gene Copy Number Amplification Promotes Rapid Evolution of Vaccinia Virus.* J Virol, 2017. **91**(4): p. e01428-16.

175. Grossegesse, M., et al., *Combined Proteomics/Genomics Approach Reveals Proteomic Changes of Mature Virions as a Novel Poxvirus Adaptation Mechanism.* Viruses, 2017. **9**(337).

176. Hatcher, E.L., Hendrickson, R.C., and Lefkowitz, E.J., *Identification of Nucleotide-Level Changes Impacting Gene Content and Genome Evolution in Orthopoxviruses.* J Virol, 2014. **88**(23): p. 13651-13668.

177. RoyChoudhury, S., Pan, A., and Mukherjee, D., *Genus specific evolution of codon usage and nucleotide compositional traits of poxviruses.* Virus Genes, 2011. **42**: p. 189-199.

178. Esteban, D.J. and Hutchinson, A.P., *Genes in the terminal regions of orthopoxvirus genomes experience adaptive molecular evolution.* BMC Genomics, 2011. **12**(261).

179. Bayer, A., Brennan, G., and Geballe, A.P., *Adaptation by copy number variation in monopartite viruses.* Curr Opin Virol, 2018. **13**(33): p. 7-12.

180. Bratke, K.A., McLysaght, A., and Rothenberg, S., *A survey of host range genes in poxvirus genomes.* Infect Genet Evol, 2013. **14**: p. 406-425.

181. Haller, S.L., et al., *Poxviruses and the evolution of host range and virulence.* Infect Genet Evol, 2014. **21**: p. 15-40.

182. Smith, G.L., et al., *Vaccinia virus immune evasion: mechanisms, virulence, and immunogenicity.* J Gen Virol, 2013. **94**(Pt 11): p. 2367-2392.

183. Qin, L., et al., *Evolution of and evolutionary relationships between extant vaccinia virus strains.* J Virol, 2015. **89**(3): p. 1809-1824.

184. Mayr, A., Hochstein-Mintzel, V., and Stickl, H., *Passage history, properties and applicability of the attenuated vaccinia virus strain MVA.* Infection, 1975. **3**(1): p. 6-14.

185. Brennan, G., et al., *Adaptive gene amplification as an intermediate step in the expansion of virus host range.* PLoS Pathog, 2014. **10**(3): p. e1004002.

186. Baldick, J., C.J. and Moss, B., *Characterization and temportal regulation of mRNAs encoded by vaccinia virus intermediate stage genes.* J Virol, 1993. **67**(6): p. 3515-3527.

187. Broyles, S.S. and Moss, B., *Homology between RNA polymerases of poxviruses, prokaryotes, and eukaryotes: nucleotide sequence and transcriptional analysis of vaccinia virus genes encoding 147-kDa and 22-kDa subunits.* Proc Natl Acad Sci USA, 1986. **83**(10): p. 3141-3145.

188. Broyles, S.S., et al., *Purification of a factor required for transcription of vaccinia virus early genes.* J Biol Chem, 1988. **263**(22): p. 10754-10760.

189. Luo, Y., Halgler, J., and Shuman, S., *Discrete functional stages of vaccinia virus early transcription during a single round of RNA synthesis in vitro.* J Biol Chem, 1991. **266**(20): p. 13303-13310.

190. Cassetti, M.C. and Moss, B., *Interaction of the 82-kDa subunit of the vaccinia virus early transcription factor heterodimer with the promoter core sequence directs downstream DNA binding of the 70-kDa subunit.* Proc Natl Acad Sci USA, 1996. **93**: p. 7540-7545.

191. Ahn, B.-Y. and Moss, B., *RNA polymerase-associated transcription specificity factor encoded by vaccinia virus.* Proc Natl Acad Sci USA, 1992. **89**(8): p. 3536-3540.

192. Senkevich, T.G., et al., *Genome sequence of a human tumorigenic poxvirus: prediction of specific host response-evasion genes.* Science, 1996. **273**(5276): p. 813-816.

193. Senkevich, T.G., et al., *The genome of molluscum contagiosum: analysis and comparison with other poxviruses.* Virology, 1997. **233**(1): p. 19-42.

194. Boyle, D.B., *Quantitative assessment of poxvirus promoters in fowlpox and vaccinia virus recombinants.* Virus Genes, 1992. **6**(3): p. 281-290.

195. Kozak, M., *Regulation of translation in eukaryotic systems.* Annu Rev Cell Biol, 1992. **8**: p. 197-225.

196. Challberg, M.D. and Englund, P.T., *Purification and properties of the deoxyribonucleic acid polymerase induced by vaccinia virus.* J Biol Chem, 1979. **254**(16): p. 7812-7819.

197. Hamilton, M.D. and Evans, D.H., *Enzymatic processing of replication and recombination intermediates by the vaccinia virus DNA polymerase.* Nucleic Acids Res, 2005. **20**(7): p. 2259-2268.

198. Czarnecki, M.W. and Traktman, P., *The vaccinia virus DNA polymerase and its processivity factor.* Virus Res, 2017. **234**: p. 193-206.

199. Falkner, F.G. and Moss, B., *Transient dominant selection of recombinant vaccinia viruses.* J Virol, 1990. **64**(6): p. 3108-3111.

200. Esposito, J., Condit, R., and Obijeski, J., *The preparation of orthopoxvirus DNA.* J Virol Methods, 1981. **2**(3): p. 174-179.

201. Simossis, V.A. and Heringa, J., *PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information.* Nucleic Acids Res, 2005. **33**(Web Server Issue, Suppl 2): p. W289-294.

202. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-1612.

203. Katsafanas, G.C. and Moss, B., *Histidine codons appended to the gene encoding the RPO22 subunit of vaccinia virus RNA polymerase facilitate the isolation and purification of functional enzyme and associated proteins from virus-infected cells.* Virology, 1999. **258**(2): p. 469-479.